Specifying the standard deviations of randomly time-varying parameters in stock assessment models based on penalized likelihood: a review of some theory and methods

Grant G. Thompson

Resource Ecology and Fisheries Management Division
Alaska Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE., Seattle, WA 98115-6349

Abstract

Most age-structured stock assessment models are currently based on a penalized likelihood approach. Specification or estimation of standard deviations for randomly time-varying parameters in such models has been described as one of the outstanding problems in contemporary stock assessment. A two-step process for producing such models is assumed here, where the standard deviations are first estimated by some approach that may or may not be based on penalized likelihood, following which those estimates are treated as known constants in a model based on penalized likelihood, regardless of the approach used to obtain those estimates. Three different likelihood "approaches" for estimating the standard deviations are considered here—penalized, marginal, and restricted—each of which produces an "estimator." Several "methods" are presented for obtaining the various estimators, including numeric maximization, iteration, recursion, closed-form solutions, Laplace approximation, and reverse-engineering. A linear-normal model is used as the example system throughout, because it is easy to comprehend and especially tractable. Consideration is given to three "cases:" multivariate with time-varying observation error variance, multivariate with constant observation error variance, and univariate with constant observation error variance. For the latter, closed-form equations for several distributions are provided, such as the probability of achieving a false positive (finding time variability where no such variability actually exists), the probability of achieving a false negative (finding no time variability where such variability actually does exist), and the relative magnitudes of the estimators conditional on achieving a false positive or not achieving a false negative. As has been recognized previously, the penalized likelihood estimator is biased downward. Here, a closed-form expression is provided for the relationship between the penalized likelihood and marginal likelihood estimators in the univariate case of the linear-

normal model with constant observation error variance. Although all of the estimators and methods are derived from the linear-normal model, each method's potential for extension to nonlinear/non-normal models is also discussed.

1. Introduction

It is standard practice for statistical age-structured assessments of marine fish stocks to allow at least some parameters, such as recruitment and the fishing mortality rate, to vary over time (Maunder and Punt 2013). Time-variability in recruitment (and sometimes fishing mortality) is often modeled as a random process arising from some statistical distribution, as distinguished from estimating each annual recruitment, or fishing mortality rate, as a free parameter. Recently, there has been increased interest in allowing other parameters to vary randomly with time as well, for example the catchability coefficient (Wilberg and Bence, 2006, Wilberg et al. 2010), selectivity parameters (Martell and Stewart 2014), the natural mortality rate (Jiao et al. 2012, Johnson et al. 2015), and growth parameters (Thorson and Minte-Vera in press). Although some alternatives have recently been developed (e.g., Gudmundsson and Gunnlaugsson 2012, Mäntyniemi et al. 2013, Nielsen and Berg 2014), the vast majority of commonly used age-structured assessment software packages estimate randomly time-varying quantities using a "penalized likelihood" approach (defined in section 3.1 below). Examples include: A-SCALA (Maunder and Watters 2003), AMAK (coded by James Ianelli http://nft.nefsc.noaa.gov/AMAK.html, Lowe et al. 2014), ASAP (Legault and Restrepo 1998, Legault 2012), BAM (Craig 2012, Williams and Shertzer 2015), CASAL (Bull et al. 2012), Coleraine (Hilborn et al. 2003, Magnusson and Hilborn 2007), ISCAM (Martell and Lima 2014, Martell and Stewart 2014), MULTIFAN-CL (Fournier et al. 1998, Hampton and Fournier 2001, Kleiber et al. 2012), and Stock Synthesis (Methot and Wetzel 2013).

The packages listed above typically require the user to specify a standard deviation (or variance, or coefficient of variation) for each randomly time-varying parameter. The main exception is the standard deviation of log-scale recruitment, which can be estimated internally in a few packages, such as BAM, CASAL, and Stock Synthesis. Either alternative (user specification or internal estimation) will cause problems if: 1) the user specifies the "wrong" value, or 2) the internal estimator is biased. One example of the former, which is frequently of concern to stock assessment review panels, is the case where the user specifies a significant amount of time-variability in a parameter that actually should exhibit none at all ("overparameterization"). With regard to the latter, *internal* estimation does not necessarily imply *unbiased* estimation, and some possible estimators may be preferable to others. Overall, specification/estimation of variances for randomly time-varying parameters has proven to be a weak point in many stock assessments, leading Maunder and Piner (2015) to list this as one of the outstanding problems in contemporary fisheries stock assessment.

The objective of this paper is to review some of the theory and methods dealing with this problem. Because terminology in this area of research can be confusing, care will be taken to show how the various estimators relate to one another, building upon previous systematizations such as that of Thorson et al. (2015). A linear-normal model (defined in section 2.2 below) will be used as the example system throughout, because it is easy to comprehend and especially tractable.

2. Preliminaries

2.1. Notation and frequently used functions

The following notational conventions will be observed:

- Functions, random variables, and integer constants are denoted by Roman letters (Table 1a); parameters and non-integer constants by Greek letters (Table 1b)
- Vectors and matrices are denoted by bold font; scalars by italic font
- Scalars and vectors are shown in lower case; matrices upper case

- Example: w_i represents element i of vector $\mathbf{w}$
- Example: $w_{i,j}$ represents element $\{i,j\}$ of matrix $\mathbf{W}$
- Example: $\mathbf{w}_j$ represents column j of matrix $\mathbf{W}$
- Example: $\mathbf{W}_i$ represents element i of a vector of matrices
- $\mathbf{I}(n)$ represents the $n\times n$ identity matrix
- $\mathbf{z}(n)$ represents the $n\times 1$ vector of zeros

Exceptions to the above conventions include commonly used function names such as ln(·), diag(·), and Γ(·), which will be written in regular (non-italic) font.

Both multivariate and univariate cases will be considered in this paper. Names and definitions of statistical function definitions for the multivariate case include the following:

- Log of multivariate normal probability density function f_{mul} for n-dimensional random variable $\mathbf{w}$ with $n\times 1$ mean vector $\boldsymbol{\mu}$ and $n\times n$ covariance matrix $\boldsymbol{\Sigma}$:

$$\ln\left(f_{mul}\left(\mathbf{w}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}\right)\right)=-\frac{n\cdot\ln(2\cdot\pi)+\ln\left(\left|\boldsymbol{\Sigma}\right|\right)+\left(\mathbf{w}-\boldsymbol{\mu}\right)'\cdot\boldsymbol{\Sigma}^{-1}\cdot\left(\mathbf{w}-\boldsymbol{\mu}\right)}{2}.$$

- Vector of row means of $n\times n$ matrix $\mathbf{W}$:

$\mathbf{m}(\mathbf{W})$, where each element i=1,2, …,n takes the form $\left(\frac{1}{n}\right)\cdot\sum_{j=1}^{n} w_{i,j}$.

- Row-wise covariance matrix of $n\times n$ matrix $\mathbf{W}$:

$\mathbf{V}_h(\mathbf{W})$, where each element $\{i=1,2,\ldots,n;\ j=1,2,\ldots,n\}$ takes the form

$$\left(\frac{1}{n-h}\right)\cdot\sum_{k=1}^{n}\left(w_{i,k}-\left(\frac{1}{n}\right)\cdot\sum_{m=1}^{n}w_{i,m}\right)\cdot\left(w_{j,k}-\left(\frac{1}{n}\right)\cdot\sum_{m=1}^{n}w_{j,m}\right),$$

and where 0 and 1 are conventional values for h.

Analogues of the above for the univariate case are as follow:

- Log of univariate normal probability density function f_{uni} for scalar random variable w with scalar mean μ and scalar variance σ^2:

$$\ln\left(f_{uni}\left(w \middle| \mu, \sigma^2 \right)\right) = -\frac{\ln(2\cdot\pi) + \ln\left(\sigma^2\right) + \frac{(w-\mu)^2}{\sigma^2}}{2} .$$

- Scalar mean of $n\times 1$ vector $\mathbf{w}$:

$$m(\mathbf{w}) = \left(\frac{1}{n}\right)\cdot \sum_{i=1}^{n} w_i .$$

- Scalar variance of $n\times 1$ vector $\mathbf{w}$:

$$v_h(\mathbf{w}) = \left(\frac{1}{n-h}\right)\cdot \sum_{i=1}^{n} \left(w_i - m(\mathbf{w})\right)^2 .$$

2.2. The linear-normal model

Although more general forms can be imagined (e.g., Laird and Ware 1982), the following version of the linear-normal model will be sufficiently general for the purposes of this paper:

- $\mathbf{X}$ is an $nfac\times nobs$ variable matrix
- $\boldsymbol{\zeta}$ is an $ndim\times 1$ constant vector
- $\boldsymbol{\Omega}$ is an $nfac\times ndim$ constant ("design") matrix
- For each k=1,2,…,$nobs$, $\mathbf{y}_k$ (a column of the $ndim\times nobs$ matrix $\mathbf{Y}$) is an $ndim\times 1$ vector related to $\mathbf{x}_k$ (a column of the $ndim\times nobs$ matrix $\mathbf{X}$) by $\mathbf{y}_k = \boldsymbol{\zeta}+\boldsymbol{\Omega}'\cdot\mathbf{x}_k$
- For each k=1,2,…,$nobs$, the observed value of $\mathbf{y}_k$, $\mathbf{yobs}_k$ (a column of the $ndim\times nobs$ matrix $\mathbf{Yobs}$), is an $ndim\times 1$ vector related to $\mathbf{y}_k$ by $\mathbf{yobs}_k = \mathbf{y}_k+\boldsymbol{\varepsilon}_k$, where $\boldsymbol{\varepsilon}_k$ (a column of the $ndim\times nobs$ matrix $\mathbf{E}$) is a multivariate normal random variable with $ndim\times 1$

mean vector $\mathbf{z}(ndim)$ and $ndim \times ndim$ covariance matrix $\Sigma\varepsilon_k$ (assumed known unless otherwise specified).

Now, suppose that the value of each $\mathbf{x}_k$ is unknown (or, worse, that even the identities of the *nfac* scalar variables comprising each vector $\mathbf{x}_k$ are unknown), or that the value of each $\mathbf{x}_k$ is known, but is simply a single realization of a large population of potential $\mathbf{x}_k$ values. In such cases, it is appropriate to view each $\mathbf{x}_k$ as a *random* vector. For simplicity, it will be assumed throughout this paper that these random vectors are multivariate normal, with constant mean vector $\mu\mathbf{x}$ and constant covariance matrix $\Sigma\mathbf{x}$, in which case the following conditions hold:

- Each $\mathbf{y}_k$ is multivariate normally distributed with mean $\mu\mathbf{y} = \zeta + \boldsymbol{\Omega}' \cdot \mu\mathbf{x}$ and covariance matrix $\Sigma\mathbf{y} = \boldsymbol{\Omega}' \cdot \Sigma\mathbf{x} \cdot \boldsymbol{\Omega}$.
- Each $\mathbf{yobs}_k$ is multivariate normally distributed with mean $\mu\mathbf{y}$ and covariance matrix $\Sigma\varepsilon_k + \Sigma\mathbf{y}$.

For each $k=1,2,\ldots,n$, a vector of deviations, δ_k (a column of the $ndim \times nobs$ matrix Δ), will be defined as the difference between $\mathbf{y}_k$ and $\mu\mathbf{y}$. Each δ_k can be thought of as a vector of *random effects*, because they are multivariate random variables, while $\mu\mathbf{y}$ can be thought of as a vector of *fixed effects*, because it is a constant (e.g., Davidian and Giltinan 2003). Because δ_k and $\mathbf{y}_k$ differ by only an additive constant, δ_k has the same covariance matrix as $\mathbf{y}_k$. Even though the covariance matrices are the same, it will be convenient to add the redundant symbol $\Sigma\delta$ ($=\Sigma\mathbf{y}$) to emphasize the fact that this matrix is independent of the fixed effects. Because the $ndim \times ndim$ matrix $\Sigma\delta$ is symmetric, it involves only $ndim \times (ndim+1)/2$ parameters. To emphasize this fact, it will typically be written as $\Sigma\delta(\sigma\delta,\rho\delta)$, where $\sigma\delta$ is an $ndim \times 1$ constant vector of standard deviations and $\rho\delta$ is an $ndim \times (ndim-1)/2$ constant vector of correlation coefficients.

2.3. Structure of the presentation

A two-step process for producing stock assessment models is assumed here, where $\sigma\delta$ and $\rho\delta$ are first estimated by some approach that may or may not be based on penalized likelihood, following which those estimates are treated as known constants in a model based on penalized likelihood, regardless of the approach used to obtain those estimates. Three different likelihood "approaches" for estimating the standard deviations are considered here—penalized, marginal, and restricted—each of which produces an "estimator." Several "methods" are presented for obtaining the various estimators, including numeric maximization, iteration, recursion, closed-form solutions, Laplace approximation, and reverse-engineering. Consideration is given to three "cases:" multivariate with time-varying observation error variance (i.e., the general case described in the preceding section), multivariate with constant observation error variance, and univariate with constant observation error variance. The approaches/estimators, methods, and cases are summarized in Table 2.

Equations and algorithms that are referenced subsequently in the document are numbered sequentially as they are introduced, except that equations and algorithms corresponding to the multivariate case with time-varying observation error variance have numbers of the form "x," equations and algorithms for the multivariate case with constant observation error variance have numbers of the form "x.1," and equations and algorithms for the univariate case with constant observation error variance is have numbers of the form "x.2."

3. Developing the likelihoods

3.1. Penalized likelihood approach

The log joint likelihood of $\mu\mathbf{y}$ and Δ, conditional on $\mathbf{Yobs}$, may be written as

$$\ln\left(lik_{jnt}\left(\boldsymbol{\mu y}, \boldsymbol{\Delta} \middle| \mathbf{Yobs}\right)\right) = \sum_{k=1}^{nobs} \ln\left(f_{mul}\left(\mathbf{yobs}_k \middle| \boldsymbol{\mu y} + \boldsymbol{\delta}_k, \boldsymbol{\Sigma\varepsilon}_k\right)\right).$$

Note that there are more parameters ($ndim \times (nobs+1)$) than data ($ndim \times nobs$) in the above, meaning that there are an infinite number of estimates that maximize the likelihood, each of which results in a perfect fit to the data. In order to obtain a unique solution, and to allow for the existence of observation error, it is necessary to constrain the parameters somehow. A common approach is to add a penalty term to the log joint likelihood, typically taking the form shown below:

$$pen\left(\boldsymbol{\Delta} \middle| \boldsymbol{\sigma\delta}, \boldsymbol{\rho\delta}\right) = \sum_{k=1}^{nobs} \ln\left(f_{mul}\left(\boldsymbol{\delta}_k \middle| \mathbf{z}(ndim), \boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta}, \boldsymbol{\rho\delta}\right)\right)\right).$$

Although exp(pen) has the form of a joint probability density function of Δ, it can be interpreted in at least two different ways: If $\sigma\delta$ and $\rho\delta$ represent parameters describing the *modeler's subjective prior belief* about the joint distribution of the elements of Δ, then exp(pen) is properly interpreted as a joint prior distribution, in which case the data (**Yobs**) should not be used to estimate the values of $\sigma\delta$ and $\rho\delta$ (i.e., $\sigma\delta$ and $\rho\delta$ are simply the constants that characterize the uncertainty associated with the modeler's subjective prior belief). More commonly, though, exp(pen) is treated as though it represents an *actual mechanism*; specifically, a stochastic process that gives rise to the elements of Δ, conditional on exp(lik_{jnt}). In this second interpretation of exp(pen), which is the interpretation that will be assumed here, it *is* proper to use the data to estimate the values of $\sigma\delta$ and $\rho\delta$. Note that exp(pen) cannot be interpreted as a likelihood, because it does not contain any data.

Summing ln(lik_{jnt}) and pen gives the penalized log likelihood:

$$\ln\left(lik_{pen}\left(\boldsymbol{\mu}\mathbf{y},\boldsymbol{\Delta},\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)=\ln\left(lik_{jnt}\left(\boldsymbol{\mu}\mathbf{y},\boldsymbol{\Delta}\,\middle|\,\mathbf{Yobs}\right)\right)+pen\left(\boldsymbol{\Delta}\,\middle|\,\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right). \quad (1)$$

The parameters to be estimated are $\boldsymbol{\mu}\mathbf{y}$, Δ, $\sigma\delta$, and $\rho\delta$.

Taking the partial derivative of Eq. (1) with respect to each element of $\boldsymbol{\mu}\mathbf{y}$ gives

$$\sum_{k=1}^{nobs}\left(\left(\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}\cdot\left(\mathbf{yobs}_k-\boldsymbol{\delta}_k-\boldsymbol{\mu}\mathbf{y}\right)\right),$$

implying that the *conditional* (on Δ) maximum likelihood estimate (MLE) of $\boldsymbol{\mu}\mathbf{y}$ when using the penalized likelihood is

$$\boldsymbol{\mu}\mathbf{y}_{pen,con}(\boldsymbol{\Delta})=\left(\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}\right)^{-1}\cdot\left(\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}\cdot\left(\mathbf{yobs}_k-\boldsymbol{\delta}_k\right)\right). \quad (2)$$

(Note that $\sigma\delta$ and $\rho\delta$ do not appear in Eq. (2).)

The vector of partial derivatives of Eq. (1) taken with respect to the elements of each δ_k is

$$\boldsymbol{\Sigma\varepsilon}_k^{-1}\cdot\left(\mathbf{yobs}_k-\boldsymbol{\mu}\mathbf{y}\right)-\left(\boldsymbol{\Sigma\varepsilon}_k^{-1}+\boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)^{-1}\right)\cdot\boldsymbol{\delta}_k,$$

implying that the *conditional* (on $\boldsymbol{\mu}\mathbf{y}$, given $\sigma\delta$ and $\rho\delta$) MLE of each δ_k when using the penalized likelihood is

$$\boldsymbol{\delta}_{pen,con}\left(\boldsymbol{\mu}\mathbf{y}\,\middle|\,\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)_k=\boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)\cdot\left(\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)\right)^{-1}\cdot\left(\mathbf{yobs}_k-\boldsymbol{\mu}\mathbf{y}\right). \quad (3)$$

(Note that, unlike Eq. (2), $\sigma\delta$ and $\rho\delta$ *do* appear in Eq. (3).)

Solving Eq. (2) and Eq. (3) simultaneously gives

$$\boldsymbol{\mu}\mathbf{y}_{pen}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)=\left(\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)+\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}\right)^{-1}\cdot\left(\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)+\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}\cdot\mathbf{yobs}_k\right) \quad (4)$$

and

$$\boldsymbol{\delta}_{pen}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)_k=\boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)\cdot\left(\boldsymbol{\Sigma\varepsilon}+\boldsymbol{\Sigma\delta}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)\right)^{-1}\cdot\left(\mathbf{yobs}_k-\boldsymbol{\mu}\mathbf{y}_{pen}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\right)\right).$$

The MLE of **y**, using the penalized likelihood, is therefore:

$$\mathbf{y}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})_k = \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\cdot\left(\boldsymbol{\Sigma\varepsilon}_k + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)^{-1}\cdot\mathbf{yobs}_k + \boldsymbol{\Sigma\varepsilon}_k\cdot\left(\boldsymbol{\Sigma\varepsilon}_k + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)^{-1}\cdot\boldsymbol{\mu}\mathbf{y}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}).$$

For a scalar multiplier ϕ, the following limiting case is of special interest:

$$\lim_{\phi\to\infty}\mathbf{y}_{pen}(\phi\cdot\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})_k = \mathbf{yobs}_k\,. \quad (5)$$

The above special case is equivalent to estimating each element of $\boldsymbol{\Delta}$ as a free parameter.

Substituting $\boldsymbol{\mu}\mathbf{y}_{pen}$ and $\Delta_{pen}(\sigma\delta, \rho\delta)$ for $\boldsymbol{\mu}\mathbf{y}$ and Δ in Eq. (1) gives the penalized log likelihood profile in $\sigma\delta$ and $\rho\delta$:

$$\ln\left(lik_{pen,pro}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,|\,\mathbf{Yobs})\right) = \sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\boldsymbol{\mu}\mathbf{y}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}),\boldsymbol{\Sigma\varepsilon}_k + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right) - \left(\frac{1}{2}\right)\cdot\left(\begin{array}{l} nobs\cdot\left(ndim\cdot\ln(2\cdot\pi) + \ln\left(\left|\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)\right) \\ + \sum_{k=1}^{nobs}\ln\left(\left|\boldsymbol{\Sigma\varepsilon}_k\right|\right) - \sum_{k=1}^{nobs}\ln\left(\left|\boldsymbol{\Sigma\varepsilon}_k + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)\end{array}\right). \quad (6)$$

<u>3.1.1. Special case: multivariate with constant observation error covariance</u>

Some of the equations pertaining to the general case remain the same for this special case except for removing the subscript k from the observation error covariance matrix. Some other equations, however, are simplified significantly. Two of these are shown below:

Eq. (4) becomes

$$\boldsymbol{\mu}\mathbf{y}_{pen} = \mathbf{m}(\mathbf{Yobs}) \quad (4.1)$$

and Eq. (6) becomes

$$\ln\left(lik_{pen,pro}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,|\,\mathbf{Yobs})\right) = \sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\mathbf{m}(\mathbf{Yobs}),\boldsymbol{\Sigma\varepsilon} + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right) - \left(\frac{nobs}{2}\right)\cdot\left(ndim\cdot\ln(2\cdot\pi) + \ln\left(\left|\boldsymbol{\Sigma\varepsilon}\right|\right) + \ln\left(\left|\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right) - \ln\left(\left|\boldsymbol{\Sigma\varepsilon} + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)\right). \quad (6.1)$$

3.1.2. Special case: univariate with constant observation error variance

In the univariate (*ndim*=1) case, **Yobs** and Δ are replaced by *nobs*×1 vectors **yobs** and **δ**, respectively; and the **Σε** and **Σδ** matrices are replaced by the scalars $\sigma\varepsilon^2$ and $\sigma\delta^2$, respectively. Then, Eq. (4) becomes

$$\mu y_{pen} = m(\mathbf{yobs}) \quad (4.2)$$

and Eq. (6) becomes

$$\ln\left(lik_{pen,pro}(\sigma\delta^2 \,|\, \mathbf{yobs})\right) = -nobs \cdot \left(\ln(2\cdot\pi) + \ln(\sigma\varepsilon) + \ln(\sigma\delta) + \frac{v_0(\mathbf{yobs})}{2\cdot\left(\sigma\varepsilon^2 + \sigma\delta^2\right)} \right). \quad (6.2)$$

3.2. Marginal likelihood approach

Unless the dimensions of the problem are very small, there are typically a large number of possible marginal likelihoods associated with the penalized likelihood, depending on which parameters are integrated out. For the purposes of this paper, the term "marginal likelihood" will be used to describe the penalized likelihood with *only* the random effects (Δ) integrated out.

Note that the penalized log likelihood (Eq. (1)) can be rewritten as follows:

$$\ln\left(lik_{pen}(\boldsymbol{\mu}\mathbf{y}, \boldsymbol{\Delta}, \boldsymbol{\sigma\delta}, \boldsymbol{\rho\delta} \,|\, \mathbf{Yobs})\right) = \sum_{k=1}^{nobs} \ln\left(f_{mul}\left(\mathbf{yobs}_k \,|\, \boldsymbol{\mu}\mathbf{y}, \boldsymbol{\Sigma\varepsilon}_k + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta}, \boldsymbol{\rho\delta})\right)\right)$$
$$+ \sum_{k=1}^{nobs} \ln\left(f_{mul}\left(\boldsymbol{\delta}_k \,\middle|\, \boldsymbol{\delta}_{pen,con}(\boldsymbol{\mu}\mathbf{y} \,|\, \boldsymbol{\sigma\delta}, \boldsymbol{\rho\delta})_k, \left(\boldsymbol{\Sigma\varepsilon}_k^{-1} + \boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta}, \boldsymbol{\rho\delta})^{-1}\right)^{-1} \right)\right).$$

In the above, note that the random effects $\boldsymbol{\delta}_k$ (*k*=1,2,…,*nobs*) do not appear anywhere in the first line of the right-hand side, and they appear only as the variables in a set of multivariate normal distributions in the second line. They can therefore be integrated out of lik_{pen}, leaving the following as the log marginal likelihood:

$$\ln\left(lik_{mar}\left(\boldsymbol{\mu}\mathbf{y},\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)=\sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\boldsymbol{\mu}\mathbf{y},\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right). \quad (7)$$

Differentiating the above with respect to the elements of $\boldsymbol{\mu}\mathbf{y}$ shows that the MLE in the penalized likelihood approach (Eq. (4)) is also the MLE in the marginal likelihood approach. The corresponding log marginal likelihood profile is thus:

$$\ln\left(lik_{mar,pro}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)=\sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\boldsymbol{\mu}\mathbf{y}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}),\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right). \quad (8)$$

3.2.1. Special case: multivariate with constant observation error covariance

The equations for this special case are the same as for the general case, except for deletion of the subscript *k* from $\boldsymbol{\Sigma\varepsilon}$ and use of Eq. (4.1) to represent the MLE of $\boldsymbol{\mu}\mathbf{y}$ rather than Eq. (4). So, $\ln(lik_{mar})$ becomes

$$\ln\left(lik_{mar}\left(\boldsymbol{\mu}\mathbf{y},\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)=\sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\boldsymbol{\mu}\mathbf{y},\boldsymbol{\Sigma\varepsilon}+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right) \quad (7.1)$$

and $\ln(lik_{mar,pro})$ becomes

$$\ln\left(lik_{mar,pro}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)=\sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\mathbf{m}(\mathbf{Yobs}),\boldsymbol{\Sigma\varepsilon}+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right). \quad (8.1)$$

3.2.2. Special case: univariate with constant observation error variance

The univariate form of $\ln(lik_{mar})$ is given by

$$\ln\left(lik_{mar}\left(\mu y,\sigma\delta\,\middle|\,\mathbf{yobs}\right)\right)=\sum_{k=1}^{nobs}\ln\left(f_{uni}\left(yobs_k\,\middle|\,\mu y,\sigma\varepsilon^2+\sigma\delta^2\right)\right) \quad (7.2)$$

and the univariate form of $\ln(lik_{mar,pro})$ is given by

$$\ln\left(lik_{mar,pro}\left(\sigma\delta\,\middle|\,\mathbf{yobs}\right)\right)=\sum_{k=1}^{nobs}\ln\left(f_{uni}\left(yobs_k\,\middle|\,m(\mathbf{yobs}),\sigma\varepsilon^2+\sigma\delta^2\right)\right). \quad (8.2)$$

3.3. Restricted likelihood approach

The restricted likelihood approach was originally formulated in terms of “error contrasts” (Patterson and Thompson 1971). However, Harville (1974, 1977) showed that a Bayesian interpretation is also possible, by imposing a uniform improper prior on the fixed effects (here, $\boldsymbol{\mu}\mathbf{y}$) and integrating them out of the marginal likelihood (see Laird and Ware (1982) for a similar exposition). Because an analogous procedure was used to obtain the marginal likelihood in Eq. (7), Harville’s interpretation will be employed here, to clarify the relationship between the marginal and restricted likelihood approaches (viz., marginal ≡ only random effects integrated out, whereas restricted ≡ mixed effects (i.e., both random *and* fixed effects) integrated out).

The log marginal likelihood (Eq. (7)) can be rewritten as follows:

$$\begin{aligned}\ln\left(lik_{mar}\left(\boldsymbol{\mu}\mathbf{y},\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)&=\left(\frac{1}{2}\right)\cdot\left(ndim\cdot\ln(2\cdot\pi)-\ln\left(\left|\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)^{-1}\right|\right)\right)\\&+\sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\boldsymbol{\mu}\mathbf{y}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}),\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right)\\&+\ln\left(f_{mul}\left(\boldsymbol{\mu}\mathbf{y}\,\middle|\,\boldsymbol{\mu}\mathbf{y}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}),\left(\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)^{-1}\right)^{-1}\right)\right).\end{aligned}$$

In the right-hand side of the above, the third line takes the form of the log of a multivariate normal distribution in $\boldsymbol{\mu}\mathbf{y}$. Because $\boldsymbol{\mu}\mathbf{y}$ does not appear anywhere in the first two lines and appears only as the variable in the third line, it can be integrated out of lik_{mar}, leaving the following:

$$\begin{aligned}\ln\left(lik_{res}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)&=\left(\frac{1}{2}\right)\cdot\left(ndim\cdot\ln(2\cdot\pi)-\ln\left(\left|\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)^{-1}\right|\right)\right)\\&+\sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\boldsymbol{\mu}\mathbf{y}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}),\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right).\quad(9)\end{aligned}$$

3.3.1. Special case: multivariate with constant observation error covariance

In this special case, $\ln(lik_{res})$ becomes

$$\ln\left(lik_{res}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)=\frac{ndim\cdot\left(\ln(2\cdot\pi)-\ln(nobs)\right)+\ln\left(\left|\boldsymbol{\Sigma\varepsilon}+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)}{2}$$
$$+\sum_{k=1}^{nobs}\ln\left(f_{mul}\left(\mathbf{yobs}_k\,\middle|\,\mathbf{m}(\mathbf{Yobs}),\boldsymbol{\Sigma\varepsilon}+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)\right). \quad (9.1)$$

3.3.2. Special case: univariate with constant observation error variance

In this special case, $\ln(lik_{res})$ becomes:

$$\ln\left(lik_{res}\left(\sigma\delta\,\middle|\,\mathbf{yobs}\right)\right)=\frac{\ln(2\cdot\pi)-\ln(nobs)+\ln\left(\sigma\varepsilon^2+\sigma\delta^2\right)}{2}$$
$$+\sum_{k=1}^{nobs}\ln\left(f_{uni}\left(yobs_k\,\middle|\,m(\mathbf{yobs}),\sigma\varepsilon^2+\sigma\delta^2\right)\right). \quad (9.2)$$

4. Maximizing the likelihoods

Except in some special cases, closed-form solutions to the MLEs for $\boldsymbol{\sigma\delta}$ and $\boldsymbol{\rho\delta}$ do not exist for any of the three objective functions (penalized log likelihood profile, Eq. (1); log marginal likelihood profile, Eq. (8); and log restricted likelihood, Eq. (9)). Numerical solutions are possible, however, at least in principle. Also, for each of the three likelihoods, an iterative algorithm for obtaining the MLE is available. The iterative algorithm for each likelihood is described in the respective subsection below, along with closed-form solutions for some special cases of particular interest.

4.1. Penalized likelihood approach

An iterative method for maximizing Eq. (6) proceeds as shown in the text box labeled "Algorithm 1." Algorithm 1 will result in an estimate regardless of whether an interior mode exists in all dimensions, but if an interior mode does not exist in some dimension(s), the corresponding element(s) in the estimate of $\sigma\delta$ will be zero. The rank of the estimate of $\Sigma\delta(\sigma\delta,\rho\delta)$ obtained by iteration will equal the number of interior modes.

4.1.1. Special case: multivariate with constant observation error covariance

Two additional approaches are available for the special case in which $\boldsymbol{\Sigma\varepsilon}$ is constant.

A recursive method for maximizing Eq. (6.1) proceeds as shown in the text box labeled “Algorithm 2.1.” If the recursive approach fails, it is because an interior mode does not exist in one or more dimensions, which causes the estimate of $\boldsymbol{\Sigma\delta}$ to be singular, meaning that the inversion on the right-hand side of the equation in step 2 cannot be computed.

At equilibrium, the equation in step 2 of the recursive approach can be rearranged to give the symmetric algebraic Riccati equation

$$\boldsymbol{\Sigma\delta}\cdot\boldsymbol{\Phi1}\cdot\boldsymbol{\Sigma\delta}+\boldsymbol{\Phi2}'\cdot\boldsymbol{\Sigma\delta}+\boldsymbol{\Sigma\delta}\cdot\boldsymbol{\Phi2}+\boldsymbol{\Phi3}=0\cdot\mathbf{I}(ndim)\,,\quad (10.1)$$

where $\boldsymbol{\Phi1}=\boldsymbol{\Sigma\varepsilon}^{-1}$, $\boldsymbol{\Phi2}=\mathbf{I}(ndim)-\boldsymbol{\Sigma\varepsilon}^{-1}\cdot\mathbf{V}_0(\mathbf{Yobs})/2$, and $\boldsymbol{\Phi3}=\boldsymbol{\Sigma\varepsilon}$. A method for solving this equation proceeds as shown in the text box labeled “Algorithm 3.1.” If an interior mode does not exist in all dimensions, some of the eigenvalues of the Hamiltonian in the Riccati approach will be imaginary. The number of dimensions for which an interior mode exists will be equal to one-half the number of real eigenvalues.

4.1.2. Special case: univariate with constant observation error variance

In this special case, the symmetric algebraic Riccati equation reduces to

$$\phi1\cdot\sigma\delta^4+2\cdot\phi2\cdot\sigma\delta^2+\phi3=0\,,\quad (10.2)$$

where $\phi1=\sigma\varepsilon^{-2}$, $\phi2=1-\sigma\varepsilon^{-2}\cdot v_0(\mathbf{yobs})/2$, and $\phi3=\sigma\varepsilon^2$. The roots of Eq. (10.2) are easier to understand in terms of the derivative of the log penalized likelihood profile (Eq. (6.2)) with respect to $\sigma\delta$:

$$\frac{-nobs\cdot\left(\sigma\delta^2+\sqrt{v_0(\mathbf{yobs})}\cdot\sigma\delta+\sigma\varepsilon^2\right)\cdot\left(\sigma\delta^2-\sqrt{v_0(\mathbf{yobs})}\cdot\sigma\delta+\sigma\varepsilon^2\right)}{\sigma\delta\cdot\left(\sigma\delta^2+\sigma\varepsilon^2\right)^2}=0\,.$$

Eq. (10.2) has either four real or four complex roots. In the event that the roots are real, two of them will be positive, which will correspond to a local minimum and local maximum of Eq. (6.2). Specifically, if $v_0(\mathbf{yobs})>4\cdot\sigma\varepsilon^2$, the penalized likelihood profile has a global maximum at 0, a local minimum at $\dfrac{\sqrt{v_0(\mathbf{yobs})}-\sqrt{v_0(\mathbf{yobs})-4\cdot\sigma\varepsilon^2}}{2}$, and a local maximum at

$$\sigma\delta_{pen}=\frac{\sqrt{v_0(\mathbf{yobs})}+\sqrt{v_0(\mathbf{yobs})-4\cdot\sigma\varepsilon^2}}{2}. \quad (11.2)\,.$$

Eq. (11.2) will be taken to represent the MLE for the penalized likelihood profile, even though it represents only a local maximum, and then only when the result is a real number. In the event that $v_0(\mathbf{yobs})<4\cdot\sigma\varepsilon^2$, the local extrema disappear, and $\sigma\delta_{pen}=0$.

4.2. Marginal likelihood approach

The iterative algorithm for maximizing Eq. (8) involves first computing a pair of matrices, the second of which is a function of the first. The first of these two matrices is the Hessian matrix corresponding to random effects (Δ) only. For the linear-normal model, this matrix, $\mathbf{\Lambda}_{ran}(\sigma\delta,\rho\delta)$, has elements $\{nobs\cdot(i-1)+k,\ nobs\cdot(j-1)+k\}$ equal to elements $\{i,j\}$ of the matrix $-\mathbf{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})^{-1}-\left(\mathbf{\Sigma\varepsilon}_k\right)^{-1}$, for all $i=1,2,\ldots,ndim$, $j=1,2,\ldots,ndim$, and $k=1,2,\ldots,nobs$; with all other elements zero. The second matrix, $\mathbf{\Sigma ave}_{ran}(\sigma\delta,\rho\delta)$, is an $ndim\times ndim$ matrix whose elements are averages, across observations, of the corresponding elements of $-\mathbf{\Lambda}_{ran}(\sigma\delta,\rho\delta)^{-1}$. That is, element $\{i,j\}$ of $\mathbf{\Sigma ave}_{ran}(\sigma\delta,\rho\delta)$ is the average, across k, of elements $\{nobs\cdot(i-1)+k,\ nobs\cdot(j-1)+k\}$ of the matrix $-\mathbf{\Lambda}_{ran}(\sigma\delta,\rho\delta)^{-1}$. The iteration proceeds as shown in the text box labeled “Algorithm 4.” In the event that the true MLE of $\mathbf{\Sigma\delta}$ is not positive-definite, this algorithm will not converge to the true MLE exactly. However, the resulting estimate will be positive-definite, and personal experience to date suggests that the resulting estimate will be close to the true MLE.

<u>4.2.1. Special case: multivariate with constant observation error covariance</u>

In this special case, the MLEs of $\sigma\delta$ and $\rho\delta$ are obtainable in closed form, because Eq. (8.1) is simply the product of multivariate normal probability density functions, all with the same variance. The MLEs can be extracted from the following equation:

$$\mathbf{\Sigma\delta}(\boldsymbol{\sigma\delta}_{mar},\boldsymbol{\rho\delta}_{mar}) = \mathbf{V}_0(\mathbf{Yobs}) - \mathbf{\Sigma\varepsilon}. \quad (12.1)$$

The above is a fairly intuitive estimator. By adding $\mathbf{\Sigma\varepsilon}$ to both sides of the equation, it implies that, in expectation, the covariance matrix of the observed values is equal to the covariance matrix of the true values (remembering that the covariance matrices of $\mathbf{Y}$ and $\mathbf{\Delta}$ are the same) plus the covariance matrix of the observation errors.

<u>4.2.2. Special case: univariate with constant observation error variance</u>

Similarly, in this special case, the MLE corresponding to Eq. (8.2) is:

$$\sigma\delta_{mar} = \sqrt{v_0(\mathbf{yobs}) - \sigma\varepsilon^2}. \quad (12.2)$$

It can be shown that $\sigma\delta_{mar}$ is always greater than $\sigma\delta_{pen}$ (as given by Eq. (11.2)) if $v_0(\mathbf{yobs}) > \sigma\varepsilon^2$.

For the univariate case with constant $\sigma\varepsilon$, one of the methods described by Methot and Taylor (2011) provides a spectrum of $\sigma\delta$ estimators, two special cases of which correspond to the penalized likelihood solution (Eq. (11.2)) and marginal likelihood solution (Eq. (12.2)) given here. Methot and Taylor suggest incorporating a “bias adjustment” term into the penalty function, which can be written, using somewhat different notation than that of the authors (and correcting a minor typographical error) as:

$$pen\left(\boldsymbol{\delta}\middle|\,\sigma\delta^2,\alpha\right)=\sum_{k=1}^{nobs}\ln\left(f_{uni}\left(\delta_k\,\middle|\,0,\sigma\delta^2\right)\right)+\frac{nobs\cdot\alpha\cdot\ln\left(\sigma\delta^2\right)}{2},$$

where α is the amount of adjustment. Three differences from the approach of Methot and Taylor should be noted: 1) The objective function used here is intended to be maximized, whereas the objective function of Methot and Taylor is intended to be minimized. 2) The bias adjustment used here (α) is equivalent to 1 minus the bias adjustment described by Methot and Taylor. 3) Although constant $\sigma\varepsilon$ is assumed for this comparison, the method of Methot and Taylor is not restricted to this special case, so their bias adjustment would normally be subscripted by year.

Including the adjustment term α in the penalty function has no impact on the penalized likelihood estimators of μy or $\boldsymbol{\delta}$, but the derivative of the bias-adjusted log penalized likelihood profile with respect to $\sigma\delta$ is (cf. the derivative shown in section 4.1.2):

$$\frac{-nobs\cdot\left(\begin{array}{l}\left(\sqrt{1-\alpha}\cdot\sigma\delta^2+\sqrt{v_0(\mathbf{yobs})}\cdot\sigma\delta+\sqrt{1-\alpha}\cdot\sigma\varepsilon^2\right)\times\\\left(\sqrt{1-\alpha}\cdot\sigma\delta^2-\sqrt{v_0(\mathbf{yobs})}\cdot\sigma\delta+\sqrt{1-\alpha}\cdot\sigma\varepsilon^2\right)\end{array}\right)}{\sigma\delta\cdot\left(\sigma\delta^2+\sigma\varepsilon^2\right)^2},$$

which, if $v_0(\mathbf{yobs})>4\cdot(1-\alpha)\cdot\sigma\varepsilon^2$, implies that the bias-adjusted penalized log likelihood profile has a local minimum at

$$\frac{\sqrt{v_0(\mathbf{yobs})}-\sqrt{v_0(\mathbf{yobs})-4\cdot(1-\alpha)\cdot\sigma\varepsilon^2}}{2\cdot\sqrt{1-\alpha}}$$

and a local maximum at

$$\frac{\sqrt{v_0(\mathbf{yobs})}+\sqrt{v_0(\mathbf{yobs})-4\cdot(1-\alpha)\cdot\sigma\varepsilon^2}}{2\cdot\sqrt{1-\alpha}}.$$

In the special case where α=0 (i.e., no bias adjustment), the local maximum corresponds to $\sigma\delta_{pen}$, and in the special case where $\alpha=\sigma\varepsilon^2/v_0(\mathbf{yobs})$, the local maximum corresponds to $\sigma\delta_{mar}$. The value of α yielding $\sigma\delta_{mar}$ is asymptotically equal to 1 minus the optimal bias adjustment reported by Methot and Taylor (for the constant-$\sigma\varepsilon$ case).

4.3. Restricted likelihood approach

The iterative algorithm for maximizing Eq. (9) is completely analogous to that used for the marginal likelihood, except that the Hessian matrix involving *mixed* effects (i.e., both $\mathbf{\Delta}$ and $\boldsymbol{\mu}\mathbf{y}$) is used instead of the Hessian matrix involving random effects (Δ) only. For the linear-normal model, the mixed-effects Hessian matrix, $\Lambda_{mix}(\sigma\delta,\rho\delta)$, is an $ndim\cdot(nobs+1)\times ndim\cdot(nobs+1)$ symmetric matrix, which can be constructed as follows:

- The upper-left $ndim\cdot nobs\times ndim\cdot nobs$ submatrix is identical to $\Lambda_{ran}(\sigma\delta,\rho\delta)$.
- Elements $\{nobs\cdot ndim+i,\ nobs\cdot(j-1)+k\}$ are equal to elements $\{i,j\}$ of the matrix $-\left(\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}$, for all i=1,2,…,$ndim$, j=1,2,…,$ndim$, and k=1,2,…,$nobs$.
- Elements $\{nobs\cdot(j-1)+k,\ nobs\cdot ndim+i\}$ are equal to elements $\{j,i\}$ of the matrix $-\left(\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}$, for all i=1,2,…,$ndim$, j=1,2,…,$ndim$, and k=1,2,…,$nobs$.
- Elements $\{nobs\cdot ndim+i,\ nobs\cdot ndim+j\}$ are equal to elements $\{i,j\}$ of the matrix $-\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}$, for all i=1,2,…,$ndim$ and j=1,2,…,$ndim$.

The restricted likelihood version of the matrix $\mathbf{\Sigma ave}_{mix}(\sigma\delta,\rho\delta)$ is constructed identically to its marginal likelihood counterpart, except that the averages are taken with respect to the elements of $-\Lambda_{mix}(\sigma\delta,\rho\delta)^{-1}$ rather than the elements of $-\Lambda_{ran}(\sigma\delta,\rho\delta)^{-1}$. The iteration proceeds as shown in the text box labeled "Algorithm 5," which is identical to Algorithm 4 except that $\mathbf{\Sigma ave}_{mix}$ is substituted for $\mathbf{\Sigma ave}_{ran}$ in step 2. As with the marginal likelihood approach, if the true MLE of $\mathbf{\Sigma\delta}$ is not positive-definite, this algorithm will not converge to the true MLE exactly. However, the resulting estimate will be positive-definite, and personal experience to date suggests that the resulting estimate will be close to the true MLE.

In the article by Methot and Taylor (2011), the third method described on page 1749 (after correcting another minor typographical error) is basically the univariate special case of Algorithm 5, except that their method uses bisection in order to speed convergence.

4.3.1. Special case: multivariate with constant observation error covariance

The MLEs corresponding to Eq. (9.1) are identical to their marginal likelihood counterparts, except that *nobs*–1 is used to scale the observation covariance rather than *nobs*:

$$\mathbf{\Sigma\delta}(\boldsymbol{\sigma\delta}_{\mathbf{res}},\boldsymbol{\rho\delta}_{\mathbf{res}}) = \mathbf{V}_1(\mathbf{Yobs}) - \mathbf{\Sigma\varepsilon}. \quad (13.1)$$

4.3.2. Special case: univariate with constant observation error variance

The MLE corresponding to Eq. (9.2) is computed similarly:

$$\sigma\delta_{res} = \sqrt{v_1(\mathbf{yobs}) - \sigma\varepsilon^2}. \quad (13.2)$$

5. Alternative formulations of the estimators

5.1. Laplace approximation

For models other than the multivariate normal model, computing the values of lik_{mar} and lik_{res} exactly can be cumbersome, because integrating out $\mathbf{\Delta}$, or $\mathbf{\Delta}$ and $\boldsymbol{\mu}\mathbf{y}$, may not yield a closed-form solution for the respective likelihood. For such models, approximations to lik_{mar} and lik_{res} based on Laplace's method have been proposed (e.g., Thorson et al. 2015). These involve adjusting the log penalized likelihood profile by a linear function of the log of the determinant of the appropriate Hessian matrix, which is then maximized numerically. Although the objective functions obtained by these methods are only approximate for non-normal models, they are exact for normal models.

5.1.1. Laplace approximation based on random effects only

For the linear-normal model, the log of the determinant of $\boldsymbol{\Lambda}_{ran}$ is

$$\ln\left(\left|\boldsymbol{\Lambda}_{ran}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)=\sum_{k=1}^{nobs}\ln\left(\left|\left(\boldsymbol{\Sigma\varepsilon}_k\right)^{-1}+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})^{-1}\right|\right),$$

and the log of the marginal likelihood profile $lik_{mar,pro}$ (Eq. (8)) can be rewritten as

$$\ln\left(lik_{mar,pro}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{Yobs}\right)\right)=\ln\left(lik_{pen,pro}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,|\,\mathbf{Yobs})\right) - \frac{\ln\left(\left|\boldsymbol{\Lambda}_{ran}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)-ndim\cdot nobs\cdot\ln(2\cdot\pi)}{2}. \quad (14)$$

5.1.2. Laplace approximation based on mixed effects

For the linear-normal model, the log of the determinant of $\boldsymbol{\Lambda}_{mix}$ is

$$\ln\left(\left|\boldsymbol{\Lambda}_{mix}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)=\ln\left(\left|\boldsymbol{\Lambda}_{ran}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)+\ln\left(\left|\sum_{k=1}^{nobs}\left(\boldsymbol{\Sigma\varepsilon}_k+\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right)^{-1}\right|\right),$$

and the log of the restricted likelihood lik_{res} (Eq. (9)) can be rewritten as:

$$\ln\left(lik_{res}\left(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,\middle|\,\mathbf{yvec}\right)\right)=\ln\left(lik_{pen,pro}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta}\,|\,\mathbf{Yobs})\right) - \frac{\ln\left(\left|\boldsymbol{\Lambda}_{mix}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right|\right)-ndim\cdot(nobs+1)\cdot\ln(2\cdot\pi)}{2}. \quad (15)$$

The objective function proposed by Thorson et al. (2015) is of the same type as Eq. (15).

5.2. Reverse-engineering the "known" variances in the univariate case with constant $\sigma\varepsilon$

For the univariate case with constant $\sigma\varepsilon$, it has so far been assumed that the value of $\sigma\varepsilon$ is known. In fact, Eqs. (11.2), (12.2), and (13.2) would seem to suggest that this assumption is necessary for estimation of $\sigma\delta_{pen}$, $\sigma\delta_{mar}$, or $\sigma\delta_{res}$. Indeed, some quantities of interest (i.e., things that might be represented by **yobs** here) lend themselves to measurement by statistically designed field experiments where the quantity is observed directly and from which estimates of $\sigma\varepsilon$ can be obtained, in which case the assumption that $\sigma\varepsilon$ is known seems reasonable. Other quantities of interest, however, such as annual deviations associated with a parameter in a selectivity function, do not, as such quantities are not observed directly and estimates of precision do not arise naturally as a result of sampling design.

In the event that the value of $\sigma\varepsilon$ is unknown, it can be reverse-engineered by exploiting the fact that, just as in the multivariate case, the value of $\sigma\delta_{pen}$ can be obtained by iteration (moreover, convergence of Algorithm 1 is easy to prove for the univariate case). Then, the value of $\sigma\varepsilon$ can be obtained from $v_0(\mathbf{yobs})$ and $\sigma\delta_{pen}$ by solving Eq. (11.2) for $\sigma\varepsilon$ as follows:

$$\sigma\varepsilon = \sqrt{\sigma\delta_{pen} \cdot \left(\sqrt{v_0(\mathbf{yobs})} - \sigma\delta_{pen}\right)}. \quad (16.2)$$

However, if a quantity is not observed directly (again, selectivity parameters are good examples), not only is it likely that $\sigma\varepsilon$ will be unknown, but it is also likely that the value of $v_0(\mathbf{yobs})$ will be unknown as well, in which case Eq. (16.2) would still not appear to solve the problem of unknown $\sigma\varepsilon$. Fortunately, the value of $v_0(\mathbf{yobs})$ itself can also be reverse-engineered, by exploiting the following facts: 1) the univariate form of Eq. (5) implies that, as $\sigma\delta$ approaches infinity, **yobs** is equal to the penalized likelihood estimate of **y**; and 2) the variance of **y** is the same as the variance of δ.

The purpose of reverse-engineering the values of $\sigma\varepsilon$ and $v_0(\mathbf{yobs})$, of course, is to enable their use in estimating $\sigma\delta_{mar}$ or $\sigma\delta_{res}$ by Eq. 12.2 or Eq 13.2 (recalling for the latter that

v_1(**yobs**)=v_0(**yobs**)·*nobs*/(*nobs*−1)). The methods proceed as shown in the text boxes labeled "Algorithm 6.2" and "Algorithm 7.2," respectively.

6. Distributions of the closed-form univariate estimators with constant $\sigma\varepsilon$

Closed-form distributions exist for many aspects of the univariate estimators in the linear-normal model with constant observation error variance. These can be derived by exploiting the fact that the sample variance of a normal random variable is proportional to a χ^2 random variable. Throughout this section, the following conventions are observed:

- The parameter β represents the ratio $\sigma\varepsilon/\sigma\delta$ (note that the denominator here is the *true* value, not the estimated value).
- The parameter c is a multiplier applied to $\sigma\varepsilon^2$ (or β^2). Although in principle c could take any real value, in practice it is restricted to the values 1 and 4.
    - For distributions related to the penalized likelihood estimator (Eq. (11.2)), c=4.
    - For distributions related to the marginal or restricted likelihood estimators (Eqs. (12.2) and (13.2)), c=1 corresponds to the case where $\sigma\varepsilon$ is known, and c=4 corresponds to the case where $\sigma\varepsilon$ is unknown, in which case Algorithm 6.2 or 7.2 is used to reverse-engineer the unknown value.
- The parameter h has the interpretation implied by the variance operator $v_h(\cdot)$.

6.1. Probability of obtaining a false positive

As noted in section 1, overparameterization is a common concern of stock assessment review panels. In the present context, overparameterization results when the estimated value of $\sigma\delta$ is positive even though the true value of $\sigma\delta$ is zero; that is, the estimate is a "false positive."

In the univariate linear-normal model with constant observation error variance, the cumulative probability of obtaining a false positive is:

$$false\_pos(c,h,nobs) = 1 - g\left(\left(c \cdot (nobs - h)\right) \,\middle|\, nobs - 1\right),$$

where $g(w|d)$ is the χ^2 cumulative distribution function evaluated at w with d degrees of freedom. Examples are shown in Figure 1. Note that the probability of obtaining a false positive is less than 0.5% for the penalized likelihood approach (c=4, h=0) at all values of *nobs*>1. In contrast, if the value of $\sigma\varepsilon$ is known, the probability of obtaining a false positive for the marginal likelihood approach (c=1, h=0) ranges from ~0.32 at *nobs*=2 to 0.5 as *nobs* approaches infinity; while the probability of obtaining a false positive for the restricted likelihood approach (c=1, h=1) ranges from ~0.16 at *nobs*=2 to 0.5 as *nobs* approaches infinity.

6.2. Probability of obtaining a false negative

The opposite problem, underparameterization, results when the estimated value of $\sigma\delta$ is zero even though the true value of $\sigma\delta$ is positive; that is, the estimate is a "false negative." The cumulative probability of obtaining a false negative is:

$$false\_neg(c,h,nobs,\beta) = g\left(c \cdot (nobs - h) \cdot \left(\frac{\beta^2}{\beta^2 + 1}\right) \,\middle|\, nobs - 1\right).$$

Examples are shown in Figure 2 (Figure 2a shows the probability as a function of β for various values of *nobs*, while Figure 2b shows the probability as a function of *nobs* for various values of β). Note that the probability of obtaining a false negative in the penalized likelihood approach can be fairly high under certain combinations of *nobs* and β values. For example, for all values of $\beta > \sqrt{1/3}$, the probability of obtaining a false negative under the penalized likelihood approach is greater than 50% for all values of *nobs*>1. On the other hand, the probability of obtaining a false negative under the marginal or restricted likelihood approaches is fairly low under a very wide range of *nobs* and β values. For all estimators, the probability of obtaining a false negative varies inversely with β (i.e., the smaller the relative value of $\sigma\varepsilon$, the less likely it is to mask the signal provided by $\sigma\delta$). For the marginal likelihood and restricted likelihood estimators, the probability of obtaining a false negative also varies inversely with *nobs*. For the penalized likelihood approach, the relationship varies inversely with *nobs* for $\beta < \sqrt{1/3}$, directly for $\beta >\sim 1.59$, and non-monotonically for $\sqrt{1/3} < \beta <\sim 1.59$.

6.3. Distribution of the estimators, given presence of a false positive

Given that there is some probability of obtaining a false positive, particularly under the marginal likelihood and restricted likelihood approaches, it is natural to wonder how large those false positive values might be (e.g., if they are likely to be extremely small, their existence is unlikely to cause much of a problem). Let u_ε represent the ratio of the estimator (either $\sigma\delta_{pen}$, $\sigma\delta_{mar}$, or $\sigma\delta_{res}$) to $\sigma\varepsilon$. Then, for i=1,2, define a probability density function of the form

$$a_i\left(u_\varepsilon \mid c,h,nobs\right)=\frac{q_i\left(u_\varepsilon\right)\cdot\left(\left(\frac{nobs-h}{2}\right)\cdot r_i\left(u_\varepsilon\right)\right)^{\frac{nobs-3}{2}}\cdot\exp\left(-\left(\frac{nobs-h}{2}\right)\cdot r_i\left(u_\varepsilon\right)\right)}{\left(\frac{1}{nobs-h}\right)\cdot\Gamma\left(\frac{nobs-1}{2}\right)\cdot false\_pos(c,h,nobs)},$$

where

$$q\left(u_\varepsilon\right)=\left[\frac{u_\varepsilon-u_\varepsilon^{-3}}{u_\varepsilon}\right] \text{ and } r\left(u_\varepsilon\right)=\left[\frac{\left(u_\varepsilon-u_\varepsilon^{-1}\right)^2}{u_\varepsilon^2+1}\right].$$

Important special cases of the functions a_1 and a_2 include the following: By setting c=4 and h=0, the function a_1 gives the distribution of $\sigma\delta_{pen}/\sigma\varepsilon$, conditional on obtaining a false positive, and has range $(1,\infty)$. The function a_2 gives the distribution of $\sigma\delta_{mar}/\sigma\varepsilon$ (obtained by setting h=0) or $\sigma\delta_{res}/\sigma\varepsilon$ (obtained by setting h=1), either with $\sigma\varepsilon$ known (obtained by setting c=1) or unknown and therefore reverse-engineered (obtained by setting c=4), conditional on obtaining a false positive, and has range $\left(\sqrt{c-1},\infty\right)$.

Figure 3a shows examples of a_1 with h=0 (upper two panels, corresponding to the penalized likelihood approach), a_2 with h=0 (middle two panels, corresponding to the marginal likelihood approach), and a_2 with h=1 (bottom two panels, corresponding to the restricted likelihood approach). The left-hand column of Figure 3a corresponds to the case in which $\sigma\varepsilon$ is unknown (c=4), and the right-hand column corresponds to the case in which $\sigma\varepsilon$ is known (c=1). Note that the results for the penalized likelihood approach do not differ for these two cases, so the top two panels are identical. The lower bound of u_ε is unity for the penalized likelihood approach, $\sqrt{3}$ for the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is unknown, and zero for the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is known.

Figure 3b shows two statistics associated with the probability density functions a_1 and a_2, both plotted as functions of *nobs*. It is important to remember that these statistics are conditional on the existence of a false positive, the probability of which varies by estimator (Figure 1) and may be very low (e.g., for the penalized likelihood estimator). The upper panel of Figure 3b shows the cumulative probability that u_ε takes a value less than or equal to unity (function a_2 only, because this cumulative probability is essentially zero for function a_1), while the lower panel shows the means of the distributions. The upper panel of Figure 3b shows that false positive estimates of $\sigma\delta$ are likely to be smaller than $\sigma\varepsilon$ for both the marginal and restricted likelihood approaches, with the probability being somewhat higher for the restricted likelihood approach. The lower panel of Figure 3b shows that the mean false positive estimate (relative to $\sigma\varepsilon$) is greater than unity for all three approaches when $\sigma\varepsilon$ is unknown, but less than unity for the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is known (the only exception being *nobs*=2 in the restricted likelihood approach, where the mean is 1.08).

6.4. Distribution of the estimators, given absence of a false negative

If a false negative is obtained, none of the estimators has a distribution, being identically zero by definition. However, it is natural to wonder about the distributions of the estimators in the event that a false negative *is not* obtained. These distributions are slightly more complicated than those for the distributions of the false positives, because they depend on β (which functions as a scale parameter) in addition to *nobs*. Let u_δ represent the ratio of the estimator (either $\sigma\delta_{pen}$, $\sigma\delta_{mar}$, or $\sigma\delta_{res}$) to the true value of $\sigma\delta$. Again letting i=1,2, define a probability density function of the form

$$b_i\left(u_\delta \mid c,h,nobs,\beta\right)=\frac{s_i\left(u_\delta \mid \beta\right)\cdot\left(\left(\frac{nobs-h}{2}\right)\cdot t_i\left(u_\delta \mid \beta\right)\right)^{\frac{nobs-3}{2}}\cdot\exp\left(-\left(\frac{nobs-h}{2}\right)\cdot t_i\left(u_\delta \mid \beta\right)\right)}{\left(\frac{1}{nobs-h}\right)\cdot\Gamma\left(\frac{nobs-1}{2}\right)\cdot\left(1-false\_neg(c,h,nobs,\beta)\right)},$$

where

$$s\left(u_\delta \mid \beta\right)=\left(\frac{\beta}{\beta^2+1}\right)\cdot q\left(\frac{u_\delta}{\beta}\right) \text{ and } t\left(u_\delta \mid \beta\right)=\left(\frac{\beta^2}{\beta^2+1}\right)\cdot r\left(\frac{u_\delta}{\beta}\right).$$

The function b_1 gives the distribution of $\sigma\delta_{pen}/\sigma\delta$ (obtained by setting c=4 and h=0), conditional on *not* obtaining a false negative, and has range (β,∞). The function b_2 gives the distribution of $\sigma\delta_{mar}/\sigma\delta$ (obtained by setting h=0) or $\sigma\delta_{res}/\sigma\delta$ (obtained by setting h=1), either with $\sigma\varepsilon$ known (obtained by setting c=1) or unknown and therefore reverse-engineered (obtained by setting c=4), again conditional on *not* obtaining a false negative, and has range $\left(\sqrt{c-1}\cdot\beta,\infty\right)$.

Figure 4a is structured similarly to Figure 3a, showing examples of b_1 (penalized likelihood approach, top two panels) and b_2 (marginal and restricted likelihood approaches, middle and bottom two panels, respectively), both when $\sigma\varepsilon$ is unknown (left column) and known (right column). The four curves in each panel of Figure 4a correspond to a 2×2 factorial design of the parameters *nobs*=10,20 and β=0.2,0.4. The lower bound of u_δ is β for the penalized likelihood approach, $\sqrt{3}\beta$ for the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is unknown, and zero for the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is known.

Figure 4b shows the cumulative probability that u_δ takes a value less than or equal to unity, plotted as a function of *nobs*, for β=0.1,0.2,0.3,0.4. The lower limit of u_δ in each panel is $\log_{10}(2)$. Figure 4c is analogous, except plotted as a function of β, for *nobs*=10,20,30,40. The upper limit of u_δ for the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is unknown is $\sqrt{1/3}$. For all parameter combinations shown, the cumulative probability is greater than 50% in in the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is known, but can drop below 50% in the penalized likelihood approach and in the marginal and restricted likelihood approaches when $\sigma\varepsilon$ is unknown if *nobs* is sufficiently low and β is sufficiently high.

Figures 4d and 4e are structured similarly to Figures 4b and 4c, except that they show the means of the respective distributions rather than cumulative probabilities. All of the estimators are asymptotically unbiased as *nobs* approaches infinity and β approaches zero. When $\sigma\varepsilon$ is unknown, all of the estimators can exhibit substantial biases (in part because low values of $\sigma\delta$

can result in false negatives), but the penalized likelihood estimator is more biased than the others except at very low values of *nobs* (e.g.,<10) or high values of β (e.g., >~0.4), as might be anticipated from Eqs. (11.2) and (12.2). When $\sigma\varepsilon$ is known, unless *nobs* is very low (e.g.,<10), the marginal likelihood estimator is biased just slightly low and the restricted likelihood estimator is essentially unbiased, whereas the penalized likelihood estimator can exhibit substantial biases.

7. Discussion

7.1. The problem of bias in the penalized likelihood approach

Fournier and Archibald (1982, especially their Eq. (5.3)) may have been the first to use a penalized likelihood approach in a fishery stock assessment context. Although a negative bias in the penalized likelihood estimate of the variance of random effects was noted in the statistics literature as early as Patterson and Thompson (1974), this bias does not seem to have been widely appreciated in the stock assessment literature for the first couple of decades or so following Fournier and Archibald (1982).

Maunder and Deriso (2003) gave the first systematic treatment of the (univariate) problem in the stock assessment literature. They presented a simulated example (their Figure 8) of a penalized *negative* log likelihood profile with a global minimum at zero, a local maximum, and a local minimum that was above the local maximum but below the marginal likelihood estimate. Apart from a change of sign, this is qualitatively identical to Eq. (6.2), which is illustrated for two example sets of parameter values in Figure 5 (*nobs* = 10, v_0(**yobs**) = 1, and $\sigma\varepsilon$ = 0.40 or 0.49). Depending on parameter values, $\sigma\delta_{pen}$ can be fairly close to $\sigma\delta_{mar}$ (e.g., it is only 13% low for the $\sigma\varepsilon$ = 0.40 example in Figure 5; best-case scenario is zero bias as the ratio $\sigma\varepsilon^2/v_0$(**yobs**) approaches zero), or the bias can be fairly substantial (e.g., 31% low for the $\sigma\varepsilon$ =

0.49 example in Figure 5; worst-case scenario is $1-\sqrt{1/3}\approx 42\%$ low as the ratio $\sigma\varepsilon^2/v_0(\mathbf{yobs})$ approaches ¼).

Whether the (biased) penalized likelihood estimator is “good enough” will depend on parameter values, on the difficulty of computing a less biased estimator such as $\sigma\delta_{mar}$ or $\sigma\delta_{res}$, and on how the assessment author or assessment reviewers perceive the relative risks of overparameterization versus underparameterization (see next section). In any case, it should be emphasized that, while the penalized likelihood estimator is biased, it is surely preferable to the common alternative of simply guessing at the appropriate value(s).

7.2. The problem of over/underparameterization

From a purely semantic perspective, there is no obvious reason to prefer overparameterization to underparameterization or vice-versa; both sound like something to be avoided, and it is not immediately obvious which is worse. In linear regression theory, there is a clear tradeoff: Overparameterized models have unbiased but imprecise estimates, while underparameterized models have biased but precise estimates; and either outcome could be preferred in a particular context. In the world of fishery stock assessment, however, the weight of opinion seems heavily tilted toward the view that overparameterization is the greater of the two evils. For example, a Google Scholar search (conducted May 12, 2015) on “stock assessment” and either “overparameterized or “overparameterization” returned 336 results, while a Google Scholar search on “stock assessment” and either “underparameterized” or “underparameterization” returned only 12 results, for a ratio of 28:1. Ludwig and Walters (1985), Walters (1997), Helu et al. (2000), Walters and Martell (2002), and Hulson and Hanselman (2014) are among those who have discussed the problem of overparameterization.

However, some recent papers have favored including more time-variability than is currently customary in stock assessments. For example, Wilberg and Bence (2006), Wilberg et al. (2010), and Martell and Stewart (2014) concluded that underparameterization can typically be expected to present greater problems than overparameterization, at least where catchability and selectivity are concerned. Nevertheless, many reviewers of stock assessments, for whatever reason(s), seem to prefer a bias toward underparameterization relative to overparameterization. When developing assessments for such reviewers, the penalized likelihood approach's small probability of obtaining a false positive and relatively high probability of obtaining a false negative may be considered advantages of that approach.

Although the distributions of the closed-form univariate estimators give less (sometimes *much* less) than a 50% chance of a false positive, the iterative algorithms for the multivariate case will always result in false positives whenever the number of parameters with random effects has been overestimated (i.e., whenever the modeler has allowed random effects to be estimated for more parameters than in the "true" model). However, personal experience to date suggests that the average estimated standard deviations for the parameters with false positives will typically be a small fraction of the average estimated standard deviations for the parameters that truly do exhibit random effects, a phenomenon which might have potential for use as a diagnostic of overparameterization (i.e., parameters whose random effects end up with very small estimated standard deviations can simply be assumed constant). Along similar lines, when applying Algorithms 4 and 5 in the multivariate case, personal experience to date suggests that slow convergence may also be a useful diagnostic of overparameterization.

7.3. The problem of nonlinearity/non-normality

Although the methods described here have all been discussed in the context of the linear-normal model, some of them may be extendable to other (nonlinear or non-normal) models (Table 2). Specifically, the methods that are most likely to be candidates for extension to other models, conditional on approach, are as follow:

- Numeric maximization: This method is probably practical for the penalized likelihood approach only. Closed-form expressions for the marginal and restricted likelihood approaches in other models are likely unobtainable. An alternative might be to derive the marginal or restricted likelihoods by the Markov chain Monte Carlo method, conditional on user-specified values for the variances of the random effects, then repeat the process enough times to obtain an accurate multivariate profile over appropriate ranges and combinations of those variances. However, this alternative would be tedious at best, unless the number of parameters exhibiting random effects is very small.
- Iteration: This method is potentially extendable for any of the three approaches. A possible concern with iterative methods in general is sensitivity to initial values, although the demonstrable convergence of Algorithm 1 for the univariate case is a hopeful sign.
- Laplace approximation: This method is potentially extendable for the marginal likelihood and restricted likelihood approaches (it does not apply to the penalized likelihood approach). In principle, Laplace approximation could be used for models with any number of dimensions. However, this is likely to prove tedious for more than 2-3 dimensions (Thorson et al. 2015).
- Reverse-engineered variances: This method is potentially extendable for the marginal likelihood and restricted likelihood approaches (use of this method for the penalized

likelihood approach is redundant, because it already involves an estimate of $\sigma\delta_{pen}$ (obtained by iteration)). Note that this method was derived specifically for the univariate case. However, it can be applied in the multivariate case by assuming $\boldsymbol{\rho\delta}=\mathbf{z}(ndim\times(ndim-1)/2)$ and by applying it to each element of the $\boldsymbol{\sigma\delta}$ vector individually. Thompson and Lauth (2012) and (e.g., Lowe et al. 2014) provide examples where this method has already been extended to nonlinear/non-normal stock assessment models.

When dealing with nonlinear/non-normal models, it might be advisable to try more than one method. Although all of the methods presented here work for the linear-normal model, some might work better than others in nonlinear/non-normal models. If one method fails, having others available to try is convenient.

As noted in section 7.1, Eq. 6.2 has the same general shape (see Figure 5) as that obtained by Maunder and Deriso (2003) for a nonlinear/non-normal model, which is a hopeful sign that at least some results for the linear-normal model may be extendable to other models.

When undertaking step #1 of Algorithms 6.2 or 7.2 in a nonlinear/non-normal model, it may take several tries to find a value of $\sigma\delta$ sufficiently high that it does not constrain the annual deviations but not so high that the solution fails to converge properly (note that as $\sigma\delta$ increases, the number of "effective" parameters increases also). It is probably best to start with a reasonably low value of $\sigma\delta$ and then increase it gradually. Because of the way that some functions are parameterized in common assessment software packages (e.g., the double normal selectivity function in Stock Synthesis, Methot and Wetzel 2013), it is also possible that one or more annual deviations may "want" to go to +/– ∞. To minimize the impact of such extreme non-normality, such “outlier” deviations should probably not be considered when making the determination that $\sigma\delta$ is no longer constraining the deviations.

7.4. The problem of assuming zero correlations

As noted in the preceding section, extension of the methods based on reverse-engineered variances to multivariate models requires assuming $\boldsymbol{\rho\delta}=\mathbf{z}(ndim\times(ndim-1)/2)$; that is, an assumption that the various vectors of deviations (in the multivariate case) are uncorrelated. In practice, this assumption is almost always made in stock assessment models. Exceptions include a pair of state-space stock assessment models: 1) Gudmundsson (1994) allowed for correlations between vectors of deviations, but noted that they will almost always have to be pre-specified; and 2) Nielsen and Berg (2014) allowed for correlations between annual vectors of age-specific fishing mortality rates, but required all of those correlations either to have the same value or to follow the process $\ln(\rho)\cdot|age_1-age_2|$ for each $\{age_1,age_2\}$ pair. Of course, assuming all correlations to be zero is parsimonious, in that it saves estimation of $ndim\times(ndim-1)/2$ parameters, which again will please stock assessment reviewers with a strong aversion to overparameterization. However, the consequences of doing so, with respect to model accuracy or predictive ability, do not appear to have been thoroughly addressed in the literature on stock assessment models based on penalized likelihood. This is an area that merits further study.

8. Acknowledgments

A negative but fascinating review of a preliminary stock assessment in 2011 by the Plan Team for the Groundfish Fisheries of the Bering Sea and Aleutian Islands and the Plan Team for the Groundfish Fisheries of the Gulf of Alaska provided the initial inspiration for developing this paper. Conversations with Jim Ianelli, Mark Maunder, Rick Methot, Hans Skaug, Ian Taylor, and Jim Thorson proved helpful along the way.

9. References

Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., Bian, R., Fu, D., 2012. CASAL (C++ algorithmic stock assessment laboratory): CASAL user manual v2.30-2012/03/21. NIWA Technical Report 135. Science Communication, NIWA, Private Bag 14901, Wellington, New Zealand.

Craig, K., 2012. The Beaufort assessment model (BAM) with application to cobia: mathematical description, implementation details, and computer code. SEDAR28-RW01. SEDAR, North Charleston, SC.

Davidian, M., Giltinan, D.M., 2003. Nonlinear models for repeated measurement data: an overview and update. J. Agric. Biol. Envir. S. 8:387-419.

Fournier, D., Archibald, C. P., 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci. 39:1195-1207.

Fournier, D.A., Hampton, J., Sibert, J.A., 1998. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. 55:2105-2116.

Gudmundsson, G., 1994. Time series analysis of catch-at-age observations. J. R. Stat. Soc. Series C Appl. Stat. 43:117-126.

Gudmundsson, G., Gunnlaugsson, T., 2012. Selection and estimation of sequential catch-at-age models. Can. J. Fish. Aquat. Sci. 69:1760-1772.

Hampton, J., Fournier, D.A., 2001. A spatially disaggregated, length-based, age-structured population model of yellowfin tuna (*Thunnus albacares*) in the western and central Pacific Ocean. Mar. Freshwater Res. 52:937-963.

Harville, D.A., 1974. Bayesian inference for variance components using only error contrasts. Biometrika 61:383-385.

Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Assoc. 72:320-338.

Helu, S.L., Sampson, D.B., Yin, Y., 2000. Application of statistical model selection criteria to the Stock Synthesis assessment program. Can. J. Fish. Aquat. Sci. 57:1784-1793.

Hilborn, R., Maunder, M., Parma, A., Ernst, B., Payne, J., Starr, P., 2003. COLERAINE: a generalized age-structured stock assessment model. User's manual version 2.0. University of Washington School of Aquatic and Fishery Sciences, PO Box 355020, Seattle, WA 98195-5020. 58 p.

Hulson, P.-J.F., Hanselman, D.H., 2014. Tradeoffs between bias, robustness, and common sense when choosing selectivity forms. Fish. Res. 158:63-73.

Jiao, Y., Smith, E.P., O'Reilly, R., Orth, D.J., 2012. Modelling non-stationary natural mortality in catch-at-age models. ICES J. Mar. Sci. 69:105-118.

Johnson, K.F., Monnahan, C.C., McGilliard, C.R., Vert-pre, K.A., Anderson, S.C., Cunningham, C.J., Hurtado-Ferro, F., Licandeo, R.R., Muradian, M.L., Ono, K., Szuwalski, C.S., Valero, J.L., Whitten, A.R., Punt, A.E., 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES J. Mar. Sci. 72:137-150.

Kleiber, P., Hampton, J., Davies, N., Hoyle, S., Fournier, D.A., 2012. MULTIFAN-CL user's guide. NOAA Inouye Research Center, NMFS/PIFSC, 1845 Wasp Blvd., Bldg. 176, Honolulu, HI 96818 USA. 152 p.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38:963-974.

Legault, C.M., 2012. Technical documentation for ASAP version 3.0. NOAA Fisheries Toolbox http://nft.nefsc.noaa.gov/. 71 p.

Legault, C.M., Restrepo, V.R., 1998. A flexible forward age-structured assessment program. ICCAT working document SCRS/98/58. Sustainable Fisheries Division Contribution SFD-98/99-16. Available from U.S. Dep. Commer., NOAA, NMFS, 75 Virginia Beach Dr., Miami, FL 33149 USA. 15 p.

Lowe, S., Ianelli, J., Palsson, W., 2014. Assessment of the Atka mackerel stock in the Bering Sea/Aleutian Islands. In Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands regions (Plan Team for the Groundfish Fisheries of the Bering Sea and Aleutian Islands, compiler), p. 1609-1691. North Pacific Fishery Management Council, 605 West 4th Ave., Suite 306, Anchorage, AK 99501. http://www.afsc.noaa.gov/REFM/Docs/2014/BSAIatka.pdf

Ludwig, D., Walters, C.J., 1985. Are age-structured models appropriate for catch-effort data? Can. J. Fish. Aquat. Sci. 42:1066-1072.

Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish Fish. 8:337-358.

Mäntyniemi, S., Uusitalo, L., Peltonen, H., Haapasaari, P., Kuikka, S., 2013. Integrated, age-structured, length-based stock assessment model with uncertain process variances, structural uncertainty, and environmental covariates: case of Central Baltic herring. Can. J. Fish. Aquat. Sci. 70:1317-1326.

Martell, S.J.D., Lima, C.W., 2014. Technical documentation for the integrated statistical catch age model. International Pacific Halibut Commission. 2320 W. Commodore Way, Seattle, WA 98199-1287 USA. 15 p.

Martell, S., Stewart, I., 2014. Towards defining good practices for modeling time-varying selectivity. Fish. Res. 158:84-95.

Maunder, M.N., Deriso, R.B., 2003. Estimation of recruitment in catch-at-age models. Can. J. Fish. Aquat. Sci. 60:1204-1216.

Maunder, M.N., Piner, K.R., 2015. Contemporary fisheries stock assessment: many issues still remain. ICES J. Mar. Sci. 72:7-18.

Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142:61-74.

Maunder, M.N., Watters, G.M., 2003. A-SCALA: An age-structured statistical catch-at-length analysis for assessing tuna stocks in the eastern Pacific Ocean. Inter.-Am. Trop. Tuna Comm. Bull. 22:435-582.

Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. 68:1744-1760.

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142:86-99.

Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fish. Res. 158:96-101.

Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545-554.

Patterson, H.D., Thompson, R., 1974. Maximum likelihood estimation of components of variance. Proceedings of the 8^{th} International Biometric Conference (L.C.A. Corsten and T. Postelnicu, eds.), p. 197-207. Editura Academiei Republicii Socialiste Rumânia.

Thompson, G.G., Lauth, R.R., 2012. Assessment of the Pacific cod stock in the Eastern Bering Sea and Aleutian Islands Area. In: Plan Team for Groundfish Fisheries of the Bering Sea/Aleutian Islands (compiler), Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands regions, p. 245-544. North Pacific Fishery Management Council, 605 West 4th Ave., Suite 306, Anchorage, AK 99501. http://www.afsc.noaa.gov/REFM/Docs/2012/BSAIpcod.pdf

Thorson, J.T., Minte-Vera, C.V., in press. Relative magnitude of cohort, age, and year effects on size at age of exploited marine fishes. Fish. Res.

Thorson, J.T., Hicks, A.C., Methot, R.D., 2015. Random effect estimation of time-varying factors in Stock Synthesis. ICES J. Mar. Sci. 72:178-185.

Walters, C., 1997. Challenges in adaptive management of riparian and coastal ecosystems. Ecol. Soc. 1:1-14.

Walters, C., Martell, S.J.D., 2002. Stock assessment needs for sustainable fisheries management. Bull. Mar. Sci. 70:629-638.

Wilberg, M.J., Bence, J.R., 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Can. J. Fish. Aquat. Sci. 63:2275-2285.

Wilberg, M.J., Thorson, J.T., Linton, B.C., Berkson, J., 2010. Incorporating time-varying catchability into population dynamic stock assessment models. Rev. Fish. Sci. 18:7-24.

Williams, E.H., Shertzer, K.W., 2015. Technical documentation of the Beaufort assessment model (BAM). NOAA Tech. Memo. NMFS-SEFSC-671. U.S. Dep. Commer., NOAA, NMFS, 101 Pivers Island Road, Beaufort, NC 28516 USA. 49 p.

Text boxes and tables

Algorithm 1:

1. Set $\boldsymbol{\sigma\delta}$ and $\boldsymbol{\rho\delta}$ equal to the values corresponding to $\mathbf{V}_0(\mathbf{Yobs})$.
2. For k=1,2,…,$nobs$, set $\delta_k=\delta_{pen}(\sigma\delta,\rho\delta)_k$.
3. Set $\boldsymbol{\sigma\delta}$ and $\boldsymbol{\rho\delta}$ equal to the values that give $\boldsymbol{\Sigma\delta}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})=\mathbf{V}_0(\Delta)$.
4. Return to step 2 and repeat until $\boldsymbol{\sigma\delta}$ and $\boldsymbol{\rho\delta}$ converge.

Algorithm 2.1:

1. Set $\boldsymbol{\Sigma\delta}$ equal to $\mathbf{V}_0(\mathbf{Yobs})$.
2. Set $\boldsymbol{\Sigma\delta} = \mathbf{V}_0(\mathbf{Yobs}) \cdot \left(\boldsymbol{\Sigma\delta}^{-1} \cdot \boldsymbol{\Sigma\varepsilon} + \mathbf{I}(ndim)\right)^{-2}$.
3. Return to step 2 and repeat until $\boldsymbol{\Sigma\delta}$ converges.
4. Set $\boldsymbol{\sigma\delta}$ and $\boldsymbol{\rho\delta}$ equal to the values corresponding to the converged value of $\boldsymbol{\Sigma\delta}$.

Algorithm 3.1:

1. Create a Hamiltonian matrix $\begin{bmatrix} \mathbf{\Phi 2} & \mathbf{\Phi 1} \\ -\mathbf{\Phi 3} & \mathbf{\Phi 2'} \end{bmatrix}$.
2. Compute the eigenvalues and the eigenvector matrix associated with the Hamiltonian.
3. Form a 2·$ndim$×$ndim$ matrix $\mathbf{\Psi}$ consisting of those columns of the eigenvector matrix that correspond to positive eigenvalues.
4. Create a matrix $\mathbf{\Psi 1}$ consisting of the first $ndim$ rows of $\mathbf{\Psi}$ and another matrix $\mathbf{\Psi 2}$ consisting of the second $ndim$ rows of $\mathbf{\Psi}$.
5. Compute $\boldsymbol{\Sigma\delta}=\mathbf{\Psi 2}\cdot\mathbf{\Psi 1}^{-1}$.
6. Set $\boldsymbol{\sigma\delta}$ and $\boldsymbol{\rho\delta}$ equal to the values corresponding to $\boldsymbol{\Sigma\delta}$.

Algorithm 4:

1. Set σδ and ρδ at the values corresponding to $\mathbf{V}_0(\mathbf{Yobs})$.
2. Set $\boldsymbol{\Sigma\delta} = \mathbf{V}_0\left(\boldsymbol{\Delta}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right) + \boldsymbol{\Sigma}\mathbf{ave}_{ran}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})$.
3. Set σδ and ρδ at the values corresponding to **Σδ**.
4. Return to step 2 and repeat until the solution converges.

Algorithm 5:

1. Set σδ and ρδ at the values corresponding to $\mathbf{V}_0(\mathbf{Yobs})$.
2. Set $\boldsymbol{\Sigma\delta} = \mathbf{V}_0\left(\boldsymbol{\Delta}_{pen}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})\right) + \boldsymbol{\Sigma}\mathbf{ave}_{mix}(\boldsymbol{\sigma\delta},\boldsymbol{\rho\delta})$.
3. Set σδ and ρδ at the values corresponding to **Σδ**.
4. Return to step 2 and repeat until the solution converges.

Algorithm 6.2:

1. Set $\sigma\delta$ at a value high enough that δ is essentially unconstrained by $\sigma\delta$.
2. Estimate δ by maximizing the penalized log likelihood conditional on $\sigma\delta$.
3. Set v_0(**yobs**) = v_0(δ).
4. Estimate $\sigma\delta_{pen}$ by the univariate special case of Algorithm 1.
5. Given these estimates of v_0(**yobs**) and $\sigma\delta_{pen}$, estimate $\sigma\varepsilon$ by Eq. (16.2).
6. Given these estimates of v_0(**yobs**) and $\sigma\varepsilon$, estimate $\sigma\delta_{mar}$ by Eq. (12.2).

Algorithm 7.2:

1. Set $\sigma\delta$ at a value high enough that δ is essentially unconstrained by $\sigma\delta$.
2. Estimate δ by maximizing the penalized log likelihood conditional on $\sigma\delta$.
3. Set v_0(**yobs**) = v_0(δ) and v_1(**yobs**) = v_1(δ)
4. Estimate $\sigma\delta_{pen}$ by the univariate special case of Algorithm 1.
5. Given these estimates of v_0(**yobs**) and $\sigma\delta_{pen}$, estimate $\sigma\varepsilon$ by Eq. (16.2).
6. Given these estimates of v_1(**yobs**) and $\sigma\varepsilon$, estimate $\sigma\delta_{res}$ by Eq. (13.2).

Table 1a. List of Roman symbols used.

Symbol	Definition
a	distribution of false positive estimates
b	distribution of estimates other than false negatives
c	multiplier applied to $\sigma\varepsilon^2$
d	degrees of freedom in χ^2 distribution
f_{mul}	multivariate normal density
f_{uni}	univariate normal density
false_neg	probability of obtaining a false negative
false_pos	probability of obtaining a false positive
g	χ^2 cumulative distribution function
h	constant used in denominator of variance
$\mathbf{I}(n)$	$n\times n$ identity matrix
i	generic index
j	generic index
k	observation index
Lik_{jnt}	joint likelihood
lik_{mar}	marginal likelihood
lik_{pen}	penalized likelihood
lik_{pro}	penalized likelihood profile
lik_{res}	restricted (residual) likelihood
$\mathbf{m}$	vector of row means of a matrix
m	scalar mean of a vector
n	generic sample size
ndim	number of dimensions
nfac	number of factors
nobs	number of observations
pen	penalty function
q	a function used in a
r	another function used in a
s	a function used in b
t	another function used in b
u_δ	ratio of univariate estimate (given a chosen estimator) to $\sigma\varepsilon$
u_ε	ratio of univariate estimate (given a chosen estimator) to $\sigma\delta$
$\mathbf{V}_h$	row-wise covariance matrix of an n-column matrix, with denominator $n-h$
v_h	scalar variance of an $n\times 1$ vector, with denominator $n-h$
w, $\mathbf{w}$, $\mathbf{W}$	generic variable
$x_{i,j}$, $\mathbf{x}_i$, $\mathbf{X}$	factor(s)
$y_{i,j}$, $\mathbf{y}_i$, $\mathbf{Y}$	randomly time-varying variable(s) of interest
$yobs_{i,j}$, $\mathbf{yobs}_i$, $\mathbf{Yobs}$	observed values of $\mathbf{Y}$
$\mathbf{yvec}$	columns of $\mathbf{Y}$ stacked in a single vector
$\mathbf{z}(n)$	$n\times 1$ vector of zeros

Table 1b. List of Greek symbols used.

Symbol	Definition
α	bias adjustment (=1–*m*(**b**) in Methot and Taylor (2011))
β	ratio of $\sigma\varepsilon$ to $\sigma\delta$
Δ	*ndim*×*nobs* matrix of time-varying deviations (process errors)
δ_k	*ndim*×1 vector of deviations in year *k* (i.e., column *k* of Δ)
ε_k	*ndim*×1 vector of observation errors in year *k*
$\mathbf{\Phi 1}$	a submatrix used to compute the Hamiltonian in the Riccati approach
$\mathbf{\Phi 2}$	a second submatrix used to compute the Hamiltonian in the Riccati approach
$\mathbf{\Phi 3}$	a third submatrix used to compute the Hamiltonian in the Riccati approach
Γ	Euler's gamma function
η	scalar multiplier
Λ_{ran}	Hessian matrix when only random effects are included
Λ_{mix}	Hessian matrix when mixed (both random and fixed) effects are included
$\mu\mathbf{y}$	mean of **y** vector
Θ	matrix used to compute vector of first (*nobs*-1)×*ndim* residuals
π	pi (3.14159…)
$\rho\delta$	vector of correlation coefficients implicit in $\Sigma\delta$
$\Sigma\delta$	covariance matrix of deviations (process errors)
$\sigma\delta$	vector of standard deviations implicit in $\Sigma\delta$
$\Sigma\varepsilon$	covariance matrix of observation errors
Ω	*nfac*×*ndim* slope matrix used to convert **x** to **y**
Ξ	matrix used to compute covariance matrix of first (*nobs*-1)×*ndim* residuals
Ψ	2·*ndim*×*ndim* matrix of eigenvectors used in the Riccati approach
$\mathbf{\Psi 1}$	first ndim rows of Ψ
$\mathbf{\Psi 2}$	second ndim rows of Ψ
ζ	*ndim*×1 intercept vector used to convert **x** to **y**

Table 2. List of methods for each of the three approaches and three cases. An arrow indicates that the algorithm or equation on the left also applies to the cases spanned by the arrow. A blank cell indicates that the method does not apply to the given approach/case. For the numeric maximization methods, the "No" entry under "Extendable to other models?" indicates that such extensions would typically be at least very tedious.

Penalized likelihood approach:

Method	Linear-normal model case: Multivariate, variable $\boldsymbol{\Sigma\varepsilon}$	Multivariate, constant $\boldsymbol{\Sigma\varepsilon}$	Univariate, constant $\sigma\varepsilon$	Extendable to other models?
Numeric maximization	Eq. (6)	Eq. (6.1)	Eq. (6.2)	Yes
Iteration	Algorithm 1	→	→	Yes
Recursion		Algorithm 2.1	→	No
Closed form		Algorithm 3.1	Eq. (11.2)	No
Laplace approximation				
Reverse-engineered variances				

Marginal likelihood approach:

Method	Linear-normal model case: Multivariate, variable $\boldsymbol{\Sigma\varepsilon}$	Multivariate, constant $\boldsymbol{\Sigma\varepsilon}$	Univariate, constant $\sigma\varepsilon$	Extendable to other models?
Numeric maximization	Eq. (7)	Eq. (7.1)	Eq. (7.2)	No
Iteration	Algorithm 4	→	→	Yes
Recursion				
Closed form		Eq. (12.1)	Eq. (12.2)	No
Laplace approximation	Eq. (14)	→	→	Yes
Reverse-engineered variances			Algorithm 6.2	Yes

Restricted likelihood approach:

Method	Linear-normal model case: Multivariate, variable $\boldsymbol{\Sigma\varepsilon}$	Multivariate, constant $\boldsymbol{\Sigma\varepsilon}$	Univariate, constant $\sigma\varepsilon$	Extendable to other models?
Numeric maximization	Eq. (8)	Eq. (8.1)	Eq. (8.2)	No
Iteration	Algorithm 5	→	→	Yes
Recursion				
Closed form		Eq. (13.1)	Eq. (13.2)	No
Laplace approximation	Eq. (15)	→	→	Yes
Reverse-engineered variances			Algorithm 7.2	Yes

Figures

Figure 1. Probability of obtaining a false positive. Blue squares: penalized, red diamonds: marginal, green triangles: restricted.

Figure 2a. Probability of obtaining a false negative as a function of β. Blue squares: *nobs*=10, red diamonds: *nobs*=20, green triangles: *nobs*=30, purple circles: *nobs*=40. Upper panel: penalized, middle panel: marginal, lower panel: restricted.

Figure 2b. Probability of obtaining a false negative as a function of *nobs*. Blue squares: β=0.1, red diamonds: β=0.2, green triangles: β=0.3, purple circles: β=0.4. Upper panel: penalized, middle panel: marginal, lower panel: restricted.

Figure 3a. Probability density functions a_1 and a_2. Blue squares: *nobs*=10, red diamonds: *nobs*=20, green triangles: *nobs*=30, purple circles: *nobs*=40. Upper panels: penalized, middle panels: marginal, lower panels: restricted. Left column: σ_ε reverse-engineered, right column: σ_ε known.

Figure 3b. Upper panel: probability of u<1 under pdfs a_1 and a_2. Lower panel: Means of pdfs a_1 and a_2. Blue squares: penalized, red diamonds: marginal, green triangles: restricted. Dashed curves: σ_ε reverse-engineered, solid curves: σ_ε known (curves for the penalized approach are the same for both σ_ε reverse-engineered and σ_ε known).

Figure 4a. Probability density functions b_1 and b_2. Blue squares: {*nobs*=10, β=0.2}, red diamonds: {*nobs*=10, β=0.4}, green triangles: {*nobs*=20, β=0.2}, purple circles: {*nobs*=20, β=0.4}. Top panels: penalized, middle panels: marginal, bottom panels: restricted. Left column: σ_ε reverse-engineered, right column: σ_ε known.

Figure 4b. Probability of u<1 in pdfs b_1 and b_2 as a function of β. Blue squares: *nobs*=10, red diamonds: *nobs*=20, green triangles: *nobs*=30, purple circles: *nobs*=40. Upper panels: penalized, middle panels: marginal, lower panels: restricted. Left column: σ_ε reverse-engineered, right column: σ_ε known.

Figure 4c. Probability of u<1 in pdfs b_1 and b_2 as a function of *nobs*. Blue squares: β=0.1, red diamonds: β=0.2, green triangles: β=0.3, purple circles: β=0.4. Upper panels: penalized, middle panels: marginal, lower panels: restricted. Left column: σ_ε reverse-engineered, right column: σ_ε known.

Figure 4d. Means of pdfs b_1 and b_2 as a function of β. Blue squares: *nobs*=10, red diamonds: *nobs*=20, green triangles: *nobs*=30, purple circles: *nobs*=40. Upper panels: penalized, middle panels: marginal, lower panels: restricted. Left column: σ_ε reverse-engineered, right column: σ_ε known.

Figure 4e. Means of pdfs b_1 and b_2 as a function of *nobs*. Blue squares: β=0.1, red diamonds: β=0.2, green triangles: β=0.3, purple circles: β=0.4. Upper panels: penalized, middle panels: marginal, lower panels: restricted. Left column: σ_ε reverse-engineered, right column: σ_ε known.

Figure 5. Two examples of the penalized log likelihood profile for the univariate case with constant $\sigma\varepsilon$. Parameter values: *nobs* = 10, $v_0(\mathbf{yobs})$ = 1, and $\sigma\varepsilon$ = 0.40 (blue curve with squares) or 0.49 (red curve with diamonds). For each value of $\sigma\varepsilon$, the dotted vertical line corresponds to the local minimum, the dashed vertical line corresponds to the local maximum ($\sigma\delta_{pen}$), and the solid vertical line corresponds to the marginal likelihood estimate ($\sigma\delta_{mar}$).

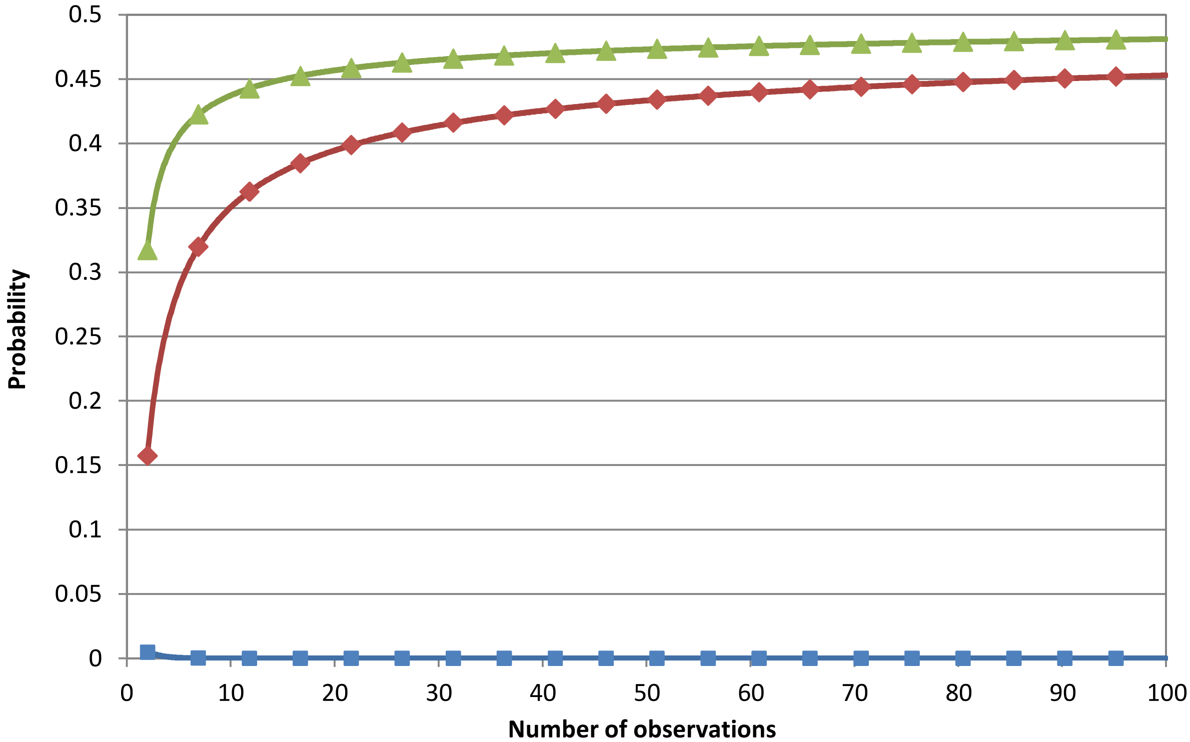

0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Probability
0
10
20
30
40
50
60
70
80
90
100
Number of observations

Figure 1.

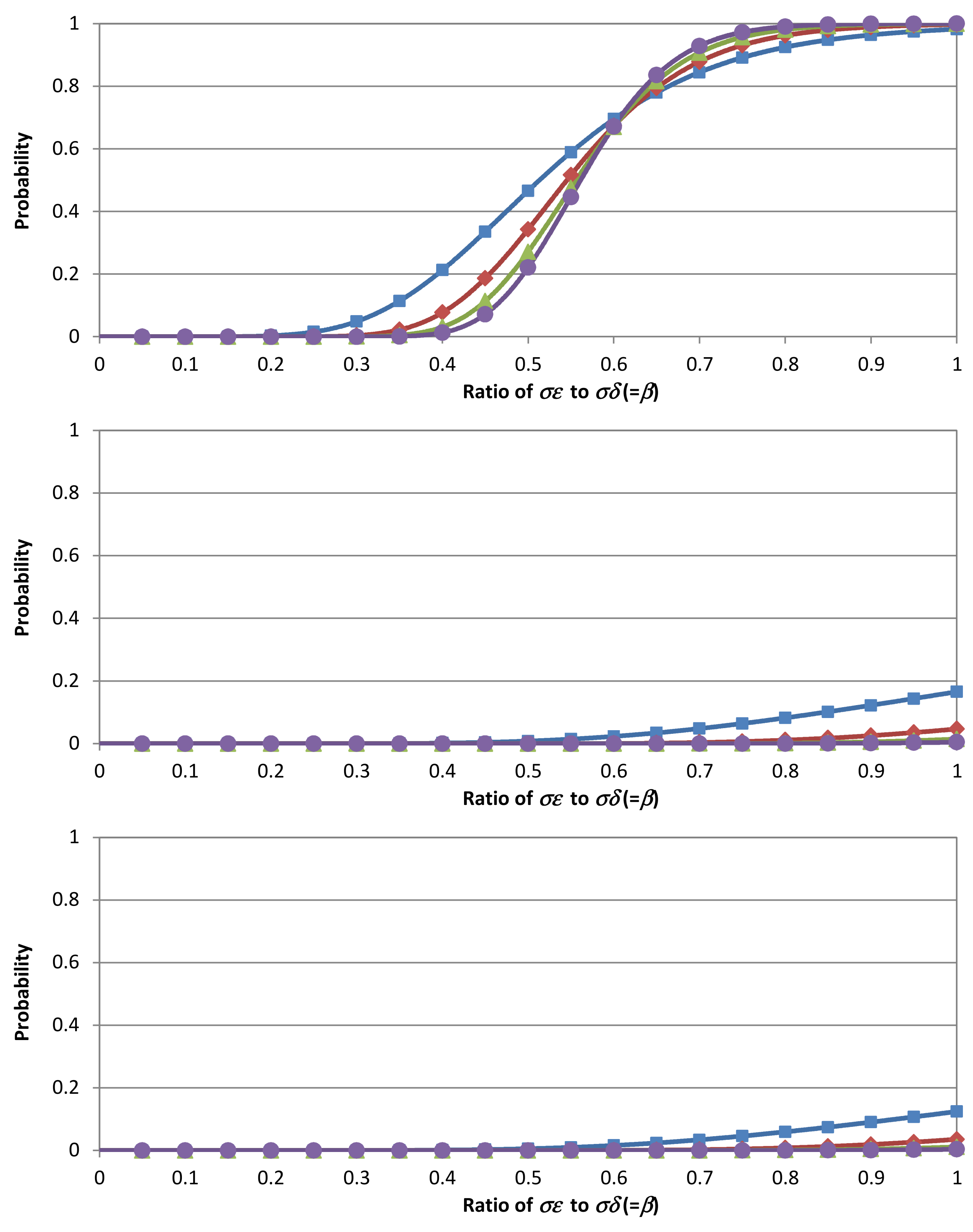

Probability
0
0.2
0.4
0.6
0.8
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Ratio of σε to σδ(=β)

Figure 2a.

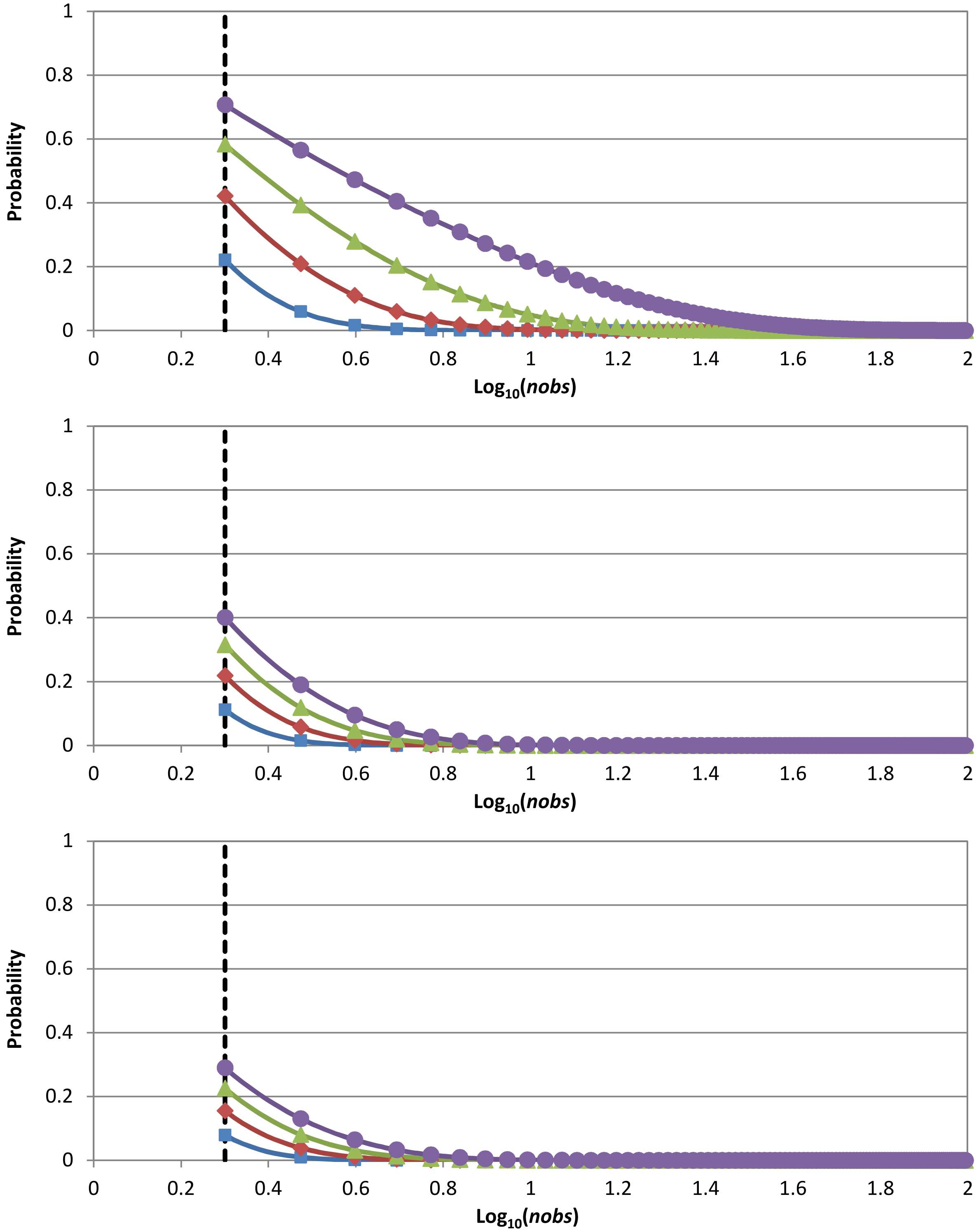


Figure 2b.

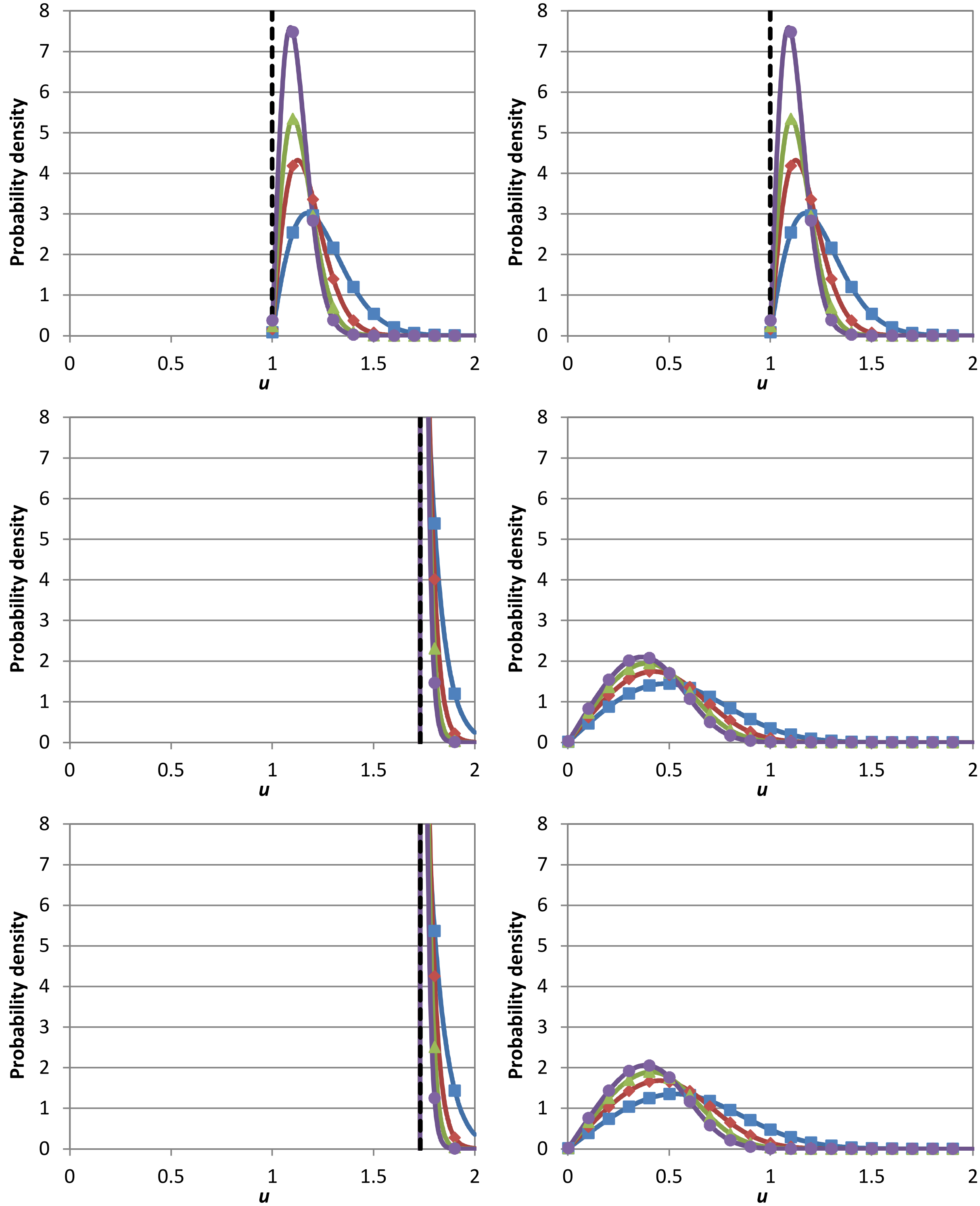
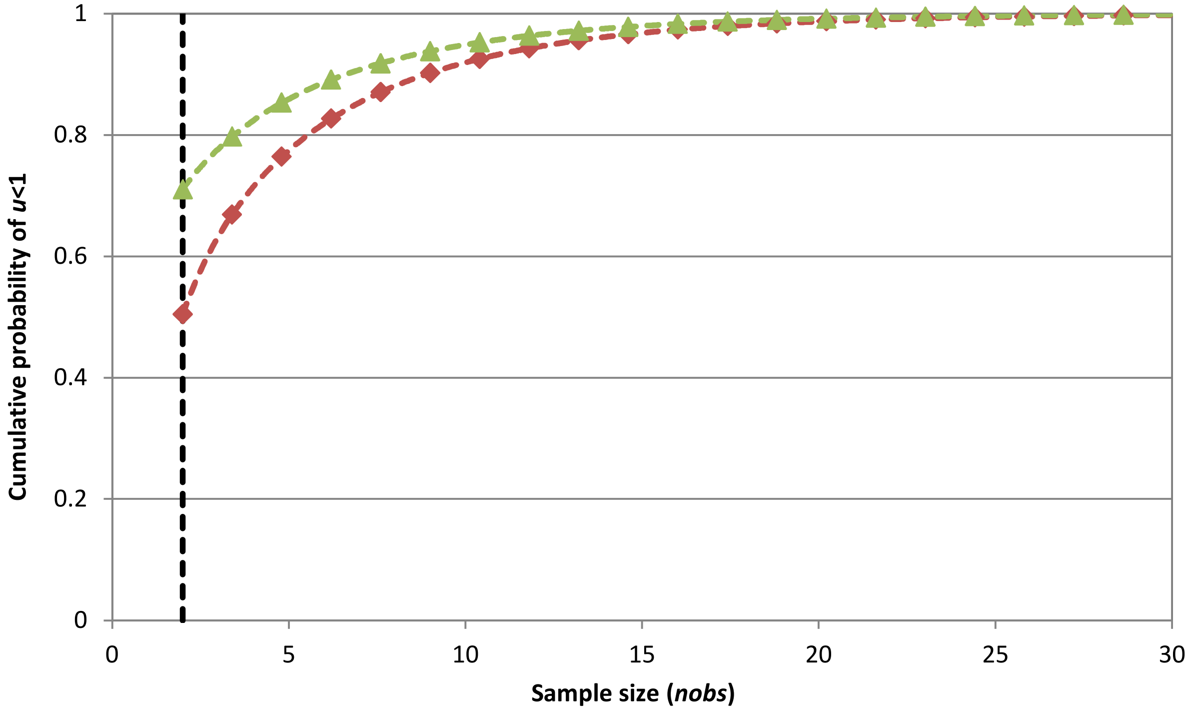

Probability density
u
Probability density
u
Probability density
u
Probability density
u
Probability density
u
Probability density
u

Figure 3a.


1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
Cumulative probability of u<1
Sample size (nobs)

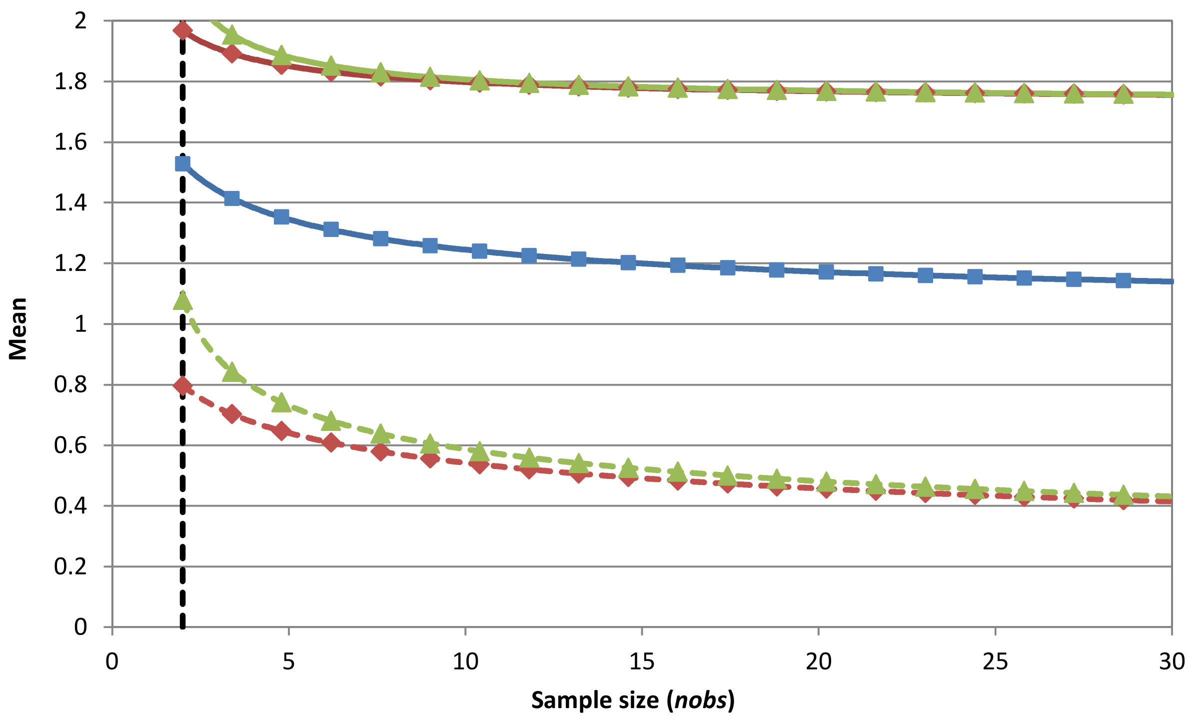

2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
30
Mean
Sample size (nobs)

Figure 3b.

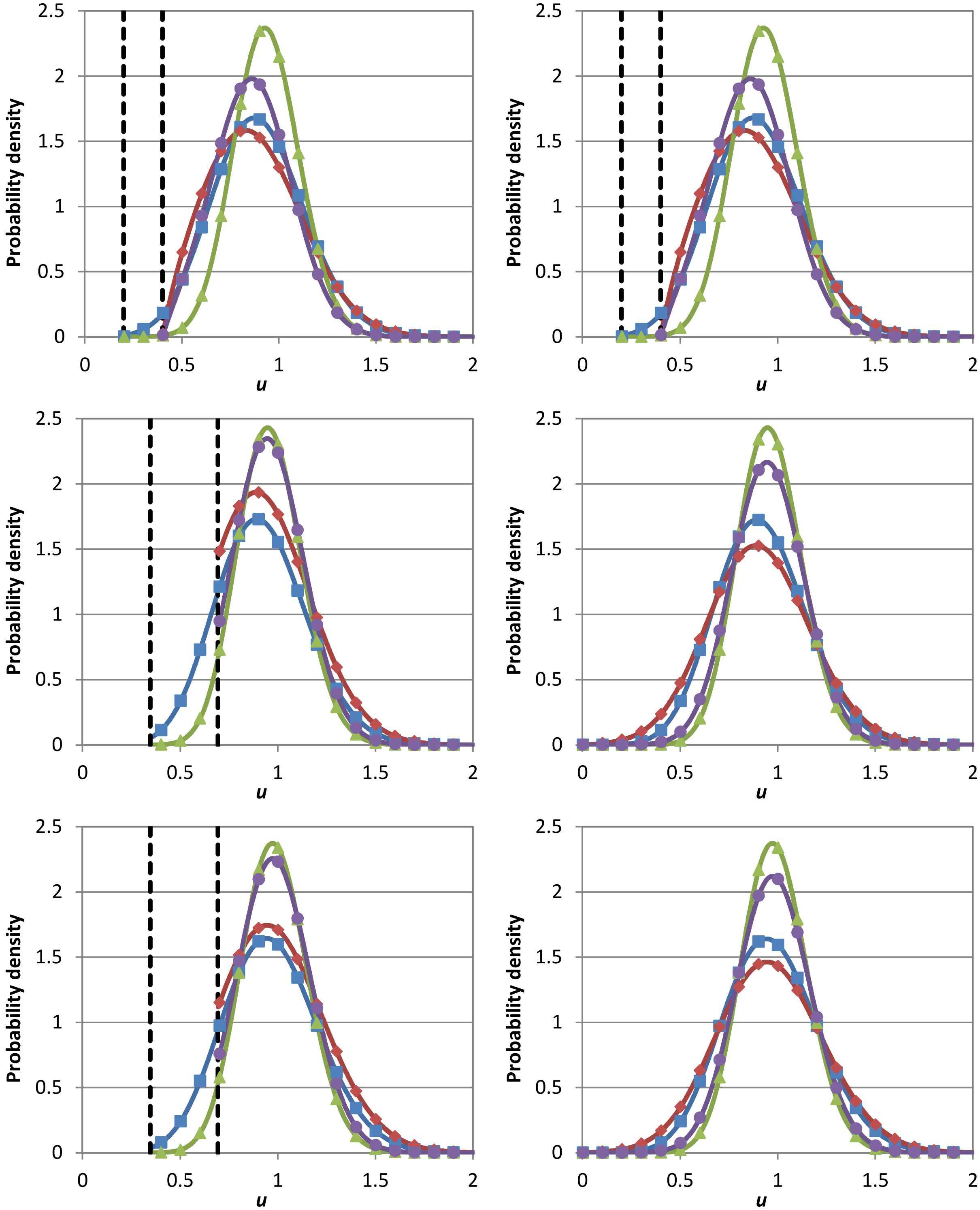


Figure 4a.

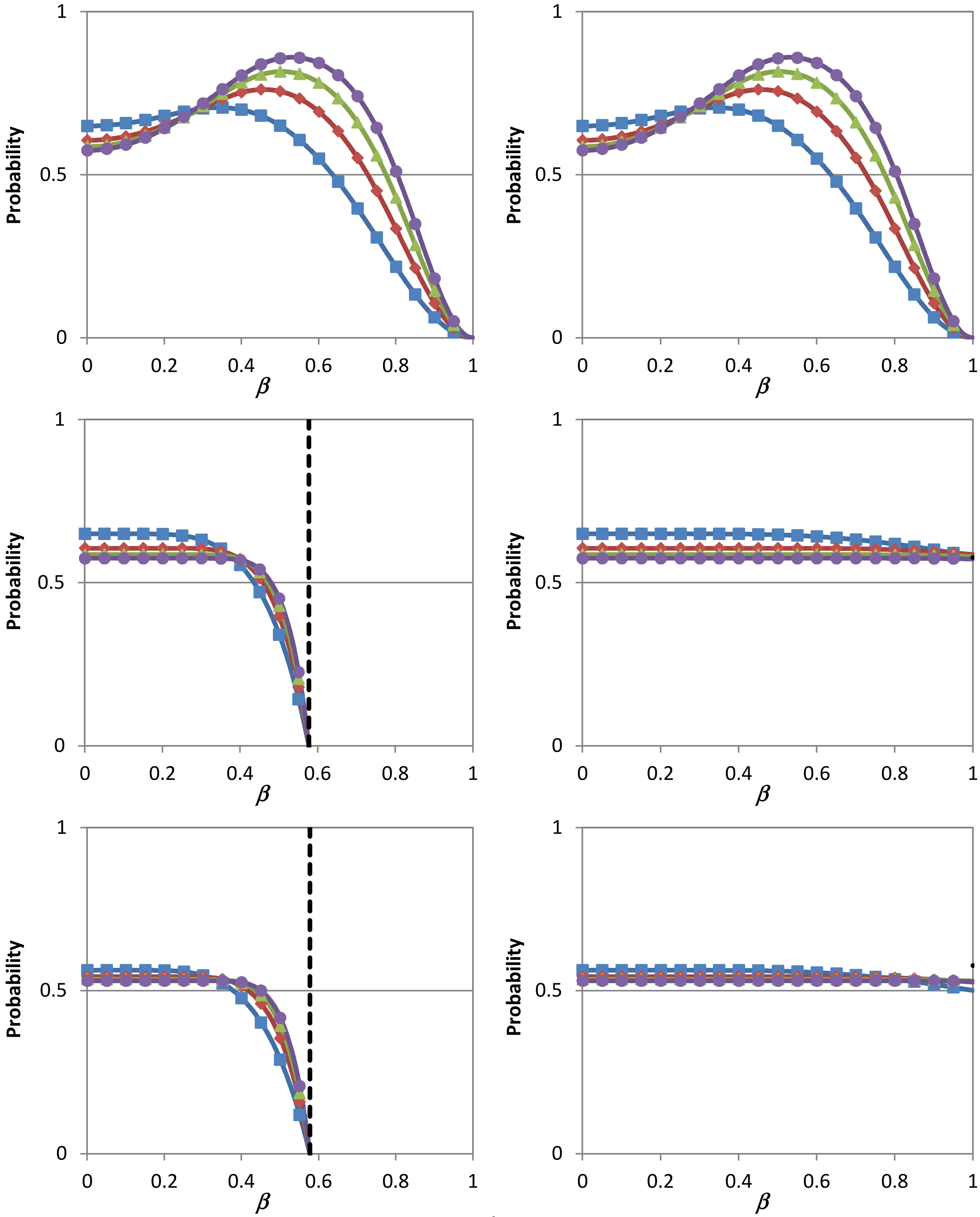
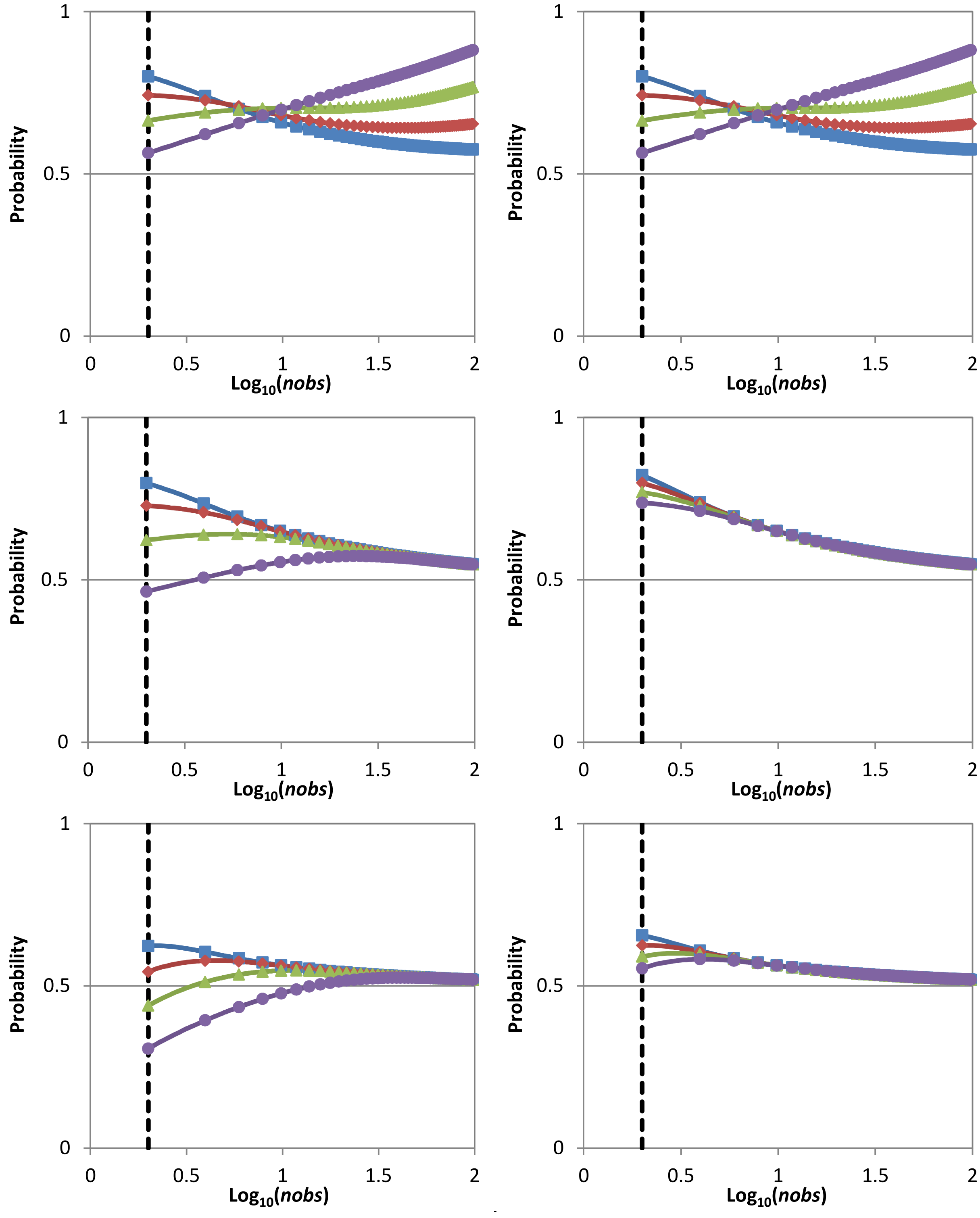
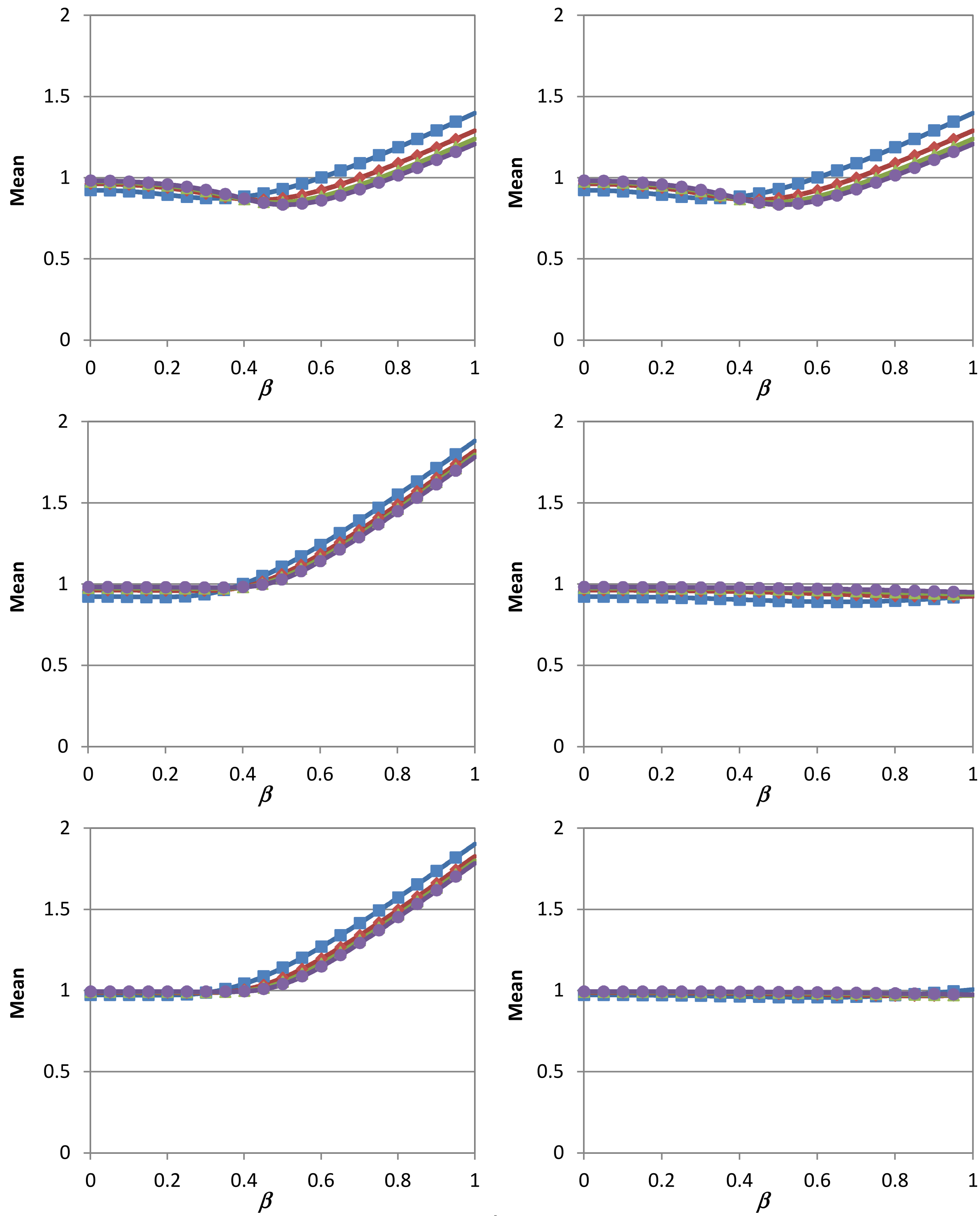


Figure 4b.



951

952 Figure 4c.



953

954 Figure 4d.

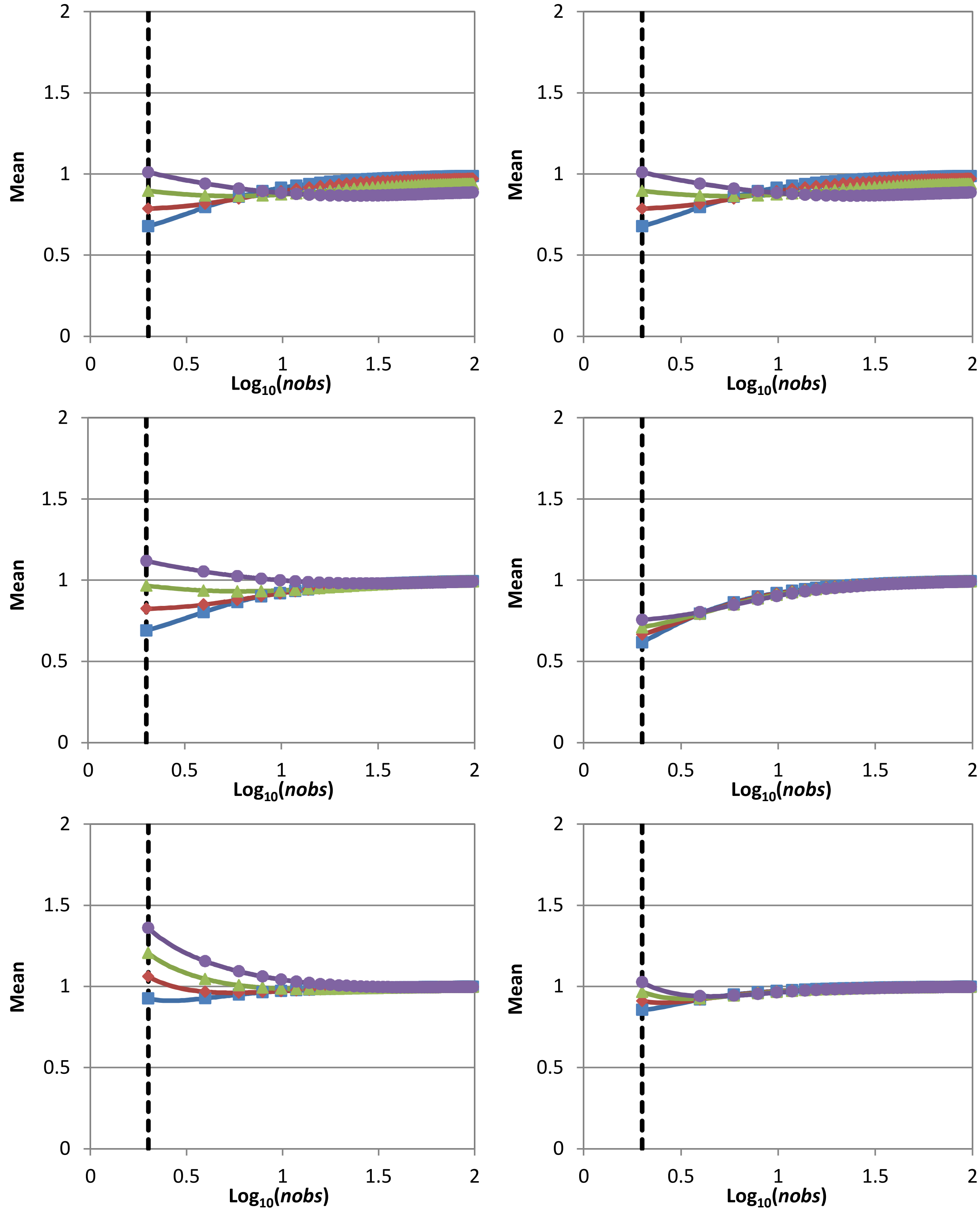


Figure 4e.

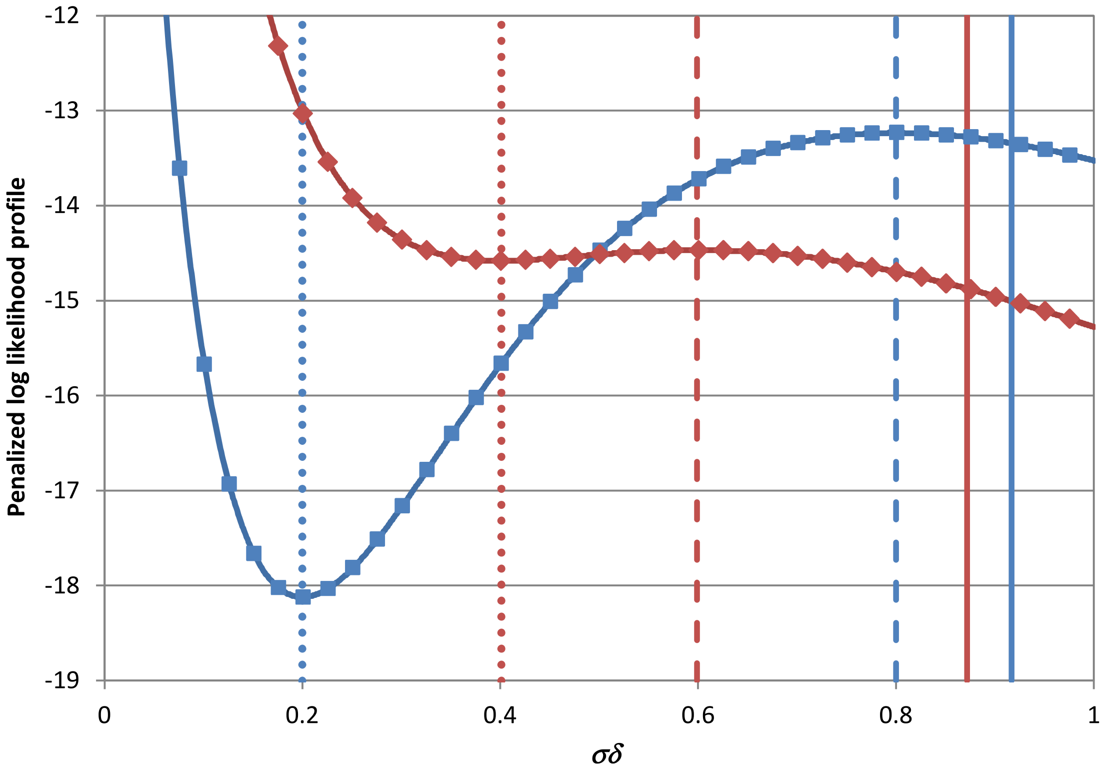

Penalized log likelihood profile
-12
-13
-14
-15
-16
-17
-18
-19
0
0.2
0.4
0.6
0.8
1
σδ

Figure 5.